

UNIVERSITÀ DEGLI STUDI DI PARMA

FACOLTÀ DI LETTERE E FILOSOFIA

CORSO DI LAUREA IN CONSERVAZIONE DEI BENI CULTURALI

IL FONDO BENUSSI CONSERVATO PRESSO LA BIBLIOTECA CENTRALE
DELL' UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA:
INVENTARIAZIONE ED IMPLEMENTAZIONE DI UNA DIGITAL LIBRARY

Relatore:

Chiar.mo Prof. ALBERTO SALARELLI

Correlatore:

Dott. MAURIZIO DI GIROLAMO

Laureanda:

SARA ADAMO

Anno Accademico 2003/2004

INDICE

INTRODUZIONE.....	p. 2
Cos'è una Digital Library?.....	p. 2
Vantaggi e svantaggi di una Digital Library.....	p. 4
Il Progetto Benussi.....	p. 6
Obiettivi del Progetto.....	p. 7
In che cosa consiste il Progetto.....	p. 7
 CAPITOLO 1: L'ARCHIVIO PERSONALE DI VITTORIO BENUSSI.....	p. 10
Chi era Benussi.....	p. 10
Notizie sull'archivio.....	p. 13
Analisi e selezione di un archivio.....	p. 16
La selezione nell'Archivio Benussi.....	p. 19
La conservazione dell'archivio cartaceo.....	p. 20
Condizioni dell'Archivio.....	p. 20
 CAPITOLO 2 : LA DIGITALIZZAZIONE.....	p. 22
In che cosa consiste la digitalizzazione.....	p. 23
Scelta dei parametri.....	p. 23
Il laboratorio di digitalizzazione.....	p. 28
La risoluzione.....	p. 31
Colore vs B/N.....	p. 34
Profondità e gradazione.....	p. 36
La compressione.....	p. 37
I formati di immagine.....	p. 39
L'ottimizzazione.....	p. 44
Le scelte per il progetto.....	p. 45
 CAPITOLO 3 : METADATI E SISTEMI DI METADATI.....	p. 50
Definizione di metadato.....	p. 50
I vari tipi di metadati.....	p. 51
Il Dublin Core.....	p. 55

L'eXtensible Mark-up Language.....	p. 59
L'Encoded Archival Description.....	p. 61
Il Resource Descriptive Format.....	p. 63
La scelta all'interno del Progetto Benussi.....	p. 66
L'UNIMARC.....	p. 68
Le ISAD(G).....	p. 71
 CAPITOLO 4 : LA GESTIONE DI IMMAGINI E METADATI.....	 p. 76
La struttura del database.....	p. 81
Il file naming system.....	p. 87
 CAPITOLO 5 :L'INTERFACCIA E L'ACCESSO.....	 p. 92
Presentazione	
Il sito.....	p. 92
L'usabilità.....	p. 92
La qualità dei siti con finalità culturali.....	p. 100
La strutturazione del sito del Progetto Benussi.....	p. 102
L'accesso.....	p. 119
Il diritto d'autore.....	p. 120
I disclaimer.....	p. 125
Le misure tecnologiche di protezione.....	p. 125
 CAPITOLO 6: SOSTENIBILITÀ E PRESERVAZIONE.....	 p. 129
La sostenibilità.....	p. 129
Preservazione.....	p. 133
Definizione.....	p. 134
Obiettivi.....	p. 135
Problemi posti dal digitale.....	p. 136
Componenti di una politica di preservazione del digitale.....	p. 139
Strategia del Progetto Benussi.....	p. 145

IN ALLEGATO: Inventario del fondo

Il tempo si affanna a consumare,
cancellare, distruggere; la
memoria cerca di prolungare
l'esistenza, o quanto meno la
vibrazione del ricordo.

(Cesare Segre)

Introduzione

Il presente lavoro consiste nell'illustrazione del percorso seguito per la creazione di una biblioteca digitale avente come scopo la conservazione e la valorizzazione di un archivio e del suo produttore, Vittorio Benussi. La digitalizzazione del fondo è il primo e più importante passo per raggiungere questo scopo, ne permette la consultazione - ovviando al problema del deterioramento del materiale - ad un numero maggiore di persone abbattendo i limiti geografici e temporali. La filosofia che è alla base di questo lavoro ben si riassume in questa frase di Stefano Vitali

Organicità nella riproduzione dei fondi, salvaguardia del contesto documentario e di riproduzione, apparati descrittivi filologicamente rigorosi sono le condizioni minime che permettono di accostarsi con consapevolezza critica alle trasposizioni digitali in formato immagine [...] degli archivi testuali.¹

Cos'è una DIGITAL LIBRARY?

Ancora oggi non esiste una determinazione univoca del concetto di biblioteca digitale in quanto caratterizzato da un complesso e da una varietà di aspetti. Alberto Salarelli afferma che nella cosiddetta "società dell'informazione" il ruolo di chi da sempre opera mediazioni informative non può che rimanere inalterato, o addirittura uscirne potenziato. La biblioteca non cambia la sua essenza, si tratta di un mutamento a livello di supporti, ma non di funzioni che essa riveste e di obiettivi che essa persegue.

...nel concetto di biblioteca digitale è possibile riconoscere una forte linea di continuità con la tradizione biblioteconomica del mondo occidentale.²

L'uso della tecnologia non cambia i concetti di base della biblioteconomia, come non cambia la funzione primaria della biblioteca, cioè quella di mettere in rapporto l'utente con le fonti documentarie.

¹ STEFANO VITALI, *Passato digitale: Le fonti dello storico nell'era del computer*, Milano, Bruno Mondadori, 2004, p. 101.

² ALBERTO SALARELLI, ANNAMARIA TAMMARO, *La biblioteca digitale*, Milano, Bibliografica, 2000, p. 76.

Al pari della scheda catalografica la funzione dei metadati è quella di fornire all'utente una serie di informazioni utili al reperimento dell'informazione potenzialmente contenuta all'interno del "digital material". Così come, al pari della scheda bibliografica, i metadati contengono informazioni relative tanto alla natura bibliografica dell'oggetto, quanto alla collocazione ove, concretamente, è possibile reperirlo (segnatura nel caso del catalogo tradizionale, *handle* nel caso del catalogo della biblioteca digitale). Senza la messa in opera delle tradizionali attività di descrizione ed indicizzazione diventa impossibile orientarsi all'interno dei materiali contenuti in una biblioteca digitale. L'universo dei documenti digitali necessita quindi di strumenti catalografici ad alto tasso di flessibilità.

La Digital Library Federation amplia ulteriormente il concetto:

Le biblioteche digitali sono organizzazioni che forniscono le risorse, compreso il personale specializzato, per selezionare, organizzare, dare l'accesso intellettuale, interpretare, distribuire, preservare l'integrità ed assicurare la persistenza nel tempo delle collezioni digitali così che queste possano essere accessibili prontamente ed economicamente per una comunità definita o per un insieme di comunità.³

La definizione di Edward Fox pone l'accento in modo ancora più marcato sul fatto che la Digital Library non è semplicemente un archivio di informazioni digitalizzate. Egli nella sua relazione ad un workshop nel 1994 scriveva:

una biblioteca digitale è un assemblaggio di archivi, informatica digitale, apparati di comunicazione con il contenuto ed il software necessari per riprodurre, emulare ed estendere i servizi forniti dalle biblioteche tradizionali cartacee, e altri mezzi per raccogliere, catalogare, individuare e distribuire l'informazione; il servizio completo di una biblioteca digitale deve realizzare tutti i servizi essenziali delle biblioteche tradizionali ed utilizzare tutti i ben noti vantaggi dell'archiviazione, della ricerca e della comunicazione digitale.⁴

Cristine Borgman divide la definizione del concetto in due parti che si rifanno alle considerazioni di due comunità, quella informatica e quella bibliotecaria:

³ La definizione è disponibile sul sito <<http://www.clir.org>>.

⁴ HENRY GLADNEY, EDWARD FOX *et al.* "Digital Library: Gross Structure and Requirements: Report from a march 1994 Workshop.", Proceedings of the First Annual Conference on the Theory and Practice of Digital Libraries, College Station - Texas, 19 – 21 giugno 1994 <<http://www.cSDL.tamu.edu/DL94/paper/fox.html>>.

1. Digital libraries are a set of electronic resources and associated technical capabilities for creating, searching, and using information. In this sense, they are an extension and enhancement of information storage and retrieval systems that manipulate digital data in any medium [...].
2. Digital libraries are constructed – collected and organized – by, and for, a community of user, and their functional capabilities support the information needs and uses of that community. They are component of communities in which individuals and groups interact with each other, using data, information and knowledge resources and system. In this sense they are extension, enhancement and integration of a variety of information institutions as physical places where resources are selected, collected, organized, preserved, and accessed in support of user community.⁵

VANTAGGI E SVANTAGGI DI UNA DIGITAL LIBRARY

L'evoluzione della biblioteca da spazio fisico a spazio virtuale comporta notevoli vantaggi, ma anche degli svantaggi, in quanto organismo nato da poco e quindi ancora imperfetto.

Il primo e più importante vantaggio è costituito dall'universalità della rappresentazione. Dal momento che ciascun medium, testo, immagine o suono è codificato in una forma unica riconducibile ad una sequenza di bit, tutte le differenti tipologie di informazione possono essere trattate nel medesimo modo e dal medesimo tipo di apparecchiatura. Inoltre le trasformazioni dell'informazione digitale sono scevre da errori mentre le trasformazioni analogiche introducono distorsioni e rumore⁶.

Annamaria Tammaro sottolinea la positività dei seguenti aspetti:

- L'accesso da parte di un pubblico vasto a una quantità assai elevata di documenti anche rari e di pregio, prima inavvicinabili. La digitalizzazione mette in atto un processo di democratizzazione in quanto rende disponibile a persone di tutte le categorie sociali, sparse in tutto il mondo, materiali rari o unici, solitamente conosciuti e raggiungibili da studiosi ed eruditi.
- La tutela di materiale in rapido degrado. La consultazione a distanza evita il maneggiamento dei supporti, uno dei fattori critici per il deterioramento.

⁵CRISTINE BORGMAN, "What are Digital Libraries? Competing visions" in *Information Processing and Management*, 38(3) 1999: 227-243.

⁶FRANCOIS FLUCKIGER, *Understanding networked multimedia: applications and technology*, London- New York, Prentice Hall, 1995, p. 21.

- La capacità di raggiungere una serie di risorse fisicamente presenti in luoghi diversi e l'opportunità di raccogliere e ordinare tali risorse in collezioni virtuali.

Nello specifico va sottolineato che al materiale digitale possono essere applicate tecniche innovative:

- Il Digital imaging, cioè la possibilità di giocare sui contrasti cromatici, sulla luminosità, sull'ingrandimento di microporzioni del supporto che consentono il recupero di una serie di dati del tutto celati all'occhio nudo o a qualsiasi, pur sofisticato, strumento analogico.
- Il restauro virtuale, cioè il ripristino del precedente status del materiale senza intervenire direttamente sul materiale stesso.

Come aspetti negativi, per cui non si è ancora riusciti a trovare una soluzione soddisfacente, si segnalano:

- I problemi derivanti dalla perdita della componente materiale del documento. La digitalizzazione non riesce a fornire l'integrità dell'informazione veicolata dal documento originale, formata dal supporto fisico e dal suo contenuto intellettuale.
- I problemi di conservazione tipici dei formati digitali. La capacità di un documento elettronico di sopravvivere nel tempo è direttamente proporzionale alla sua facile riproducibilità, ed è inversamente proporzionale alla labilità dei supporti finora utilizzati. L'informazione digitale è estremamente deperibile a causa del deterioramento fisico dei supporti e dell'obsolescenza tecnologica dei diversi prodotti digitali. Gregory e Morelli esprimono tale perplessità asserendo:

Pensando ai libri, ai documenti cartacei, alle tele dipinte, esiste una cultura di osservazione sulla durata di tali materiali lungo secoli e secoli. Le pellicole fotografiche, e a maggior ragione i supporti magnetici, sono invece appena nati. Quanto tempo durerà un CD?⁷

⁷TULLIO GREGORY, MARCELLO MORELLI (a cura di), *L'eclisse delle memorie*, Roma – Bari, Laterza, 1994.

La vita media di un cd-rom va dai cinque ai venticinque anni; mentre hardware, software e sistemi operativi hanno cicli di vita dai due ai cinque anni. Se a fronte dell'evoluzione tecnologica degli strumenti di lettura e dei supporti, elementi che mettono a rischio la conservazione, la soluzione più comoda, più pratica, più rapida ed efficace consiste nella replicazione della traccia digitale, nessuna soluzione può essere messa in atto contro l'obsolescenza degli strumenti di decodifica dei dati, se non il periodico trasferimento delle memorie digitali dai vecchi ai nuovi supporti.

- I rischi derivanti da inadeguati procedimenti di conversione. I rischi maggiori sono quelli che derivano da imperfetti o inadeguati procedimenti di digitalizzazione degli originali analogici: quando la digitalizzazione è inappropriata, i dati risultano inevitabilmente modificati e se ne rende più difficile il recupero in altro formato.

IL PROGETTO

L'idea di realizzare questo progetto è nata nel 2001, allorché alcuni docenti del Dipartimento di Psicologia dell'Università degli studi di Milano - Bicocca hanno deciso di valorizzare, con la creazione di una Digital Library, l'archivio di Vittorio Benussi loro pervenuto grazie al lascito da parte del professor Cesare Musatti, docente presso l'Istituto di Psicologia dell'Università Statale di Milano.

Nel 2002 è stata presentata richiesta per un cofinanziamento al Ministero dell'Istruzione dell'Università e della Ricerca per un progetto interuniversitario in collaborazione con l'Università degli studi di Bologna e Trieste per l'area 11, quella delle Scienze storiche, filosofiche, pedagogiche e psicologiche. Il MIUR ha assegnato al progetto di ricerca 90.000 euro, mentre L'Università degli studi di Milano ha destinato altri 42.500 euro per un totale di 132,500 euro.

Il coordinatore del progetto, il professor Natale Stucchi, ordinario di Psicologia generale presso la Facoltà di Psicologia, è stato affiancato per l'esegesi e lo studio dei documenti da un equipe di storici della scienza: il professor Mauro Antonelli, all'epoca docente dell'Ateneo di Bologna, attualmente professore associato di Storia della Psicologia presso l'Università di Milano Bicocca; la

professoressa Serena Cattaruzza, prima ricercatrice di Filosofia del linguaggio ed ora docente di Filosofia teoretica presso l'Università degli Studi di Trieste; la dott.ssa Barbara Marino, assegnista di ricerca dell'Università degli studi di Milano Bicocca, e la dott.ssa Verena Zudini, assegnista di ricerca dell'Università degli studi di Milano Bicocca.

La parte tecnica è stata curata dal personale della Biblioteca d'Ateneo: Sara Adamo si è occupata della progettazione in collaborazione con il prof. Stucchi e con il dott. Maurizio Di Girolamo, direttore della Biblioteca d'Ateneo dell'Università di Milano Bicocca, dell'ordinamento e della descrizione dell'archivio in collaborazione con il prof. Stucchi, con il prof. Antonelli e con la dott. Zudini, dell'ottimizzazione delle immagini, della creazione dei record archivistici e della realizzazione del sito web; Mariagrazia Pistelli e Luigi Zeliani, system managers di Aleph, hanno adattato Aleph alle esigenze del progetto. Fabrizio Palermo, dottorando della Facoltà di Informatica dell'Università di Milano Bicocca, ha collaborato al progetto personalizzando l'opac di Aleph e realizzando alcune componenti del sito web.

GLI OBIETTIVI

Come già detto in precedenza, due sono essenzialmente gli obiettivi del progetto: la conservazione dell'archivio e la valorizzazione dei documenti, in esso contenuti, e del suo produttore.

Il fine della conservazione viene raggiunto attraverso la digitalizzazione dei documenti in quanto il surrogato digitale evita la consultazione dell'originale e di conseguenza il facile deperimento del suo supporto.

Il fine della valorizzazione viene invece raggiunto in vari modi. Innanzitutto la creazione di una versione digitale, facilmente collocabile sul web, permette la consultazione ad un numero maggiore di persone, abbattendo i limiti geografici e temporali, e la visibilità a documenti finora rimasti inediti. La creazione di strumenti di corredo facilita la consultazione del materiale presente all'interno dell'archivio. Infine la realizzazione di un apparato critico permette un approccio ipertestuale ai documenti ed alle nozioni in essi presenti.

IN CHE COSA CONSISTE IL PROGETTO

La realizzazione della suddetta biblioteca digitale si è articolata in 5 fasi.

- 1) Tra il marzo 2002 e luglio 2002, dopo la lettura di documentazione tecnica sull'argomento, sono state progettate tutte le procedure, che sarebbero poi state messe in atto durante il progetto, ed acquistate le attrezzature hardware e software necessarie. Tra l'agosto 2002 ed il febbraio 2003 l'archivio è stato preparato per la digitalizzazione; tale preparazione è consistita nell'ordinamento di quei fascicoli che nelle passate consultazioni erano rimasti fuori posto e nella descrizione del fondo archivistico in base alle ISAD(G), cioè l'International Standard of Archival Description; infine è stata compiuta una selezione del materiale che doveva essere o no digitalizzato.
- 2) Nella seconda fase, svoltasi nei mesi di marzo e aprile 2003, è stata realizzata la digitalizzazione del materiale archivistico attraverso la scansione dei documenti. Il lavoro è stato eseguito da una società di servizi di Milano, la Gallo Pomi srl, in outsourcing anche se fisicamente il lavoro è stato compiuto nei locali della biblioteca. Tra giugno e novembre 2003 è stato svolto il lavoro di ottimizzazione delle immagini, consistente nel controllo dei file, nella scontornatura⁸ e nella creazione di immagini in formato compresso per la visualizzazione tramite il web.
- 3) Tra marzo e luglio 2003 è stata attuata anche la terza fase del progetto, la creazione dei metadati per mezzo del software di gestione già in possesso della biblioteca; il gestionale utilizza per la descrizione bibliografica il formato UNIMARC, che è stato adattato alle necessità di descrizione dell'archivio e del progetto. Dopo la fase di personalizzazione si è proceduto alla creazione dei record di descrizione.
- 4) La fase successiva ha richiesto l'ideazione di meccanismi per l'accesso e la loro concreta creazione. La procedura di ricerca e di accesso alle descrizioni viene compiuta dal software gestionale, a cui non sono stati

⁸ Il termine è usato nella "Normativa per l'acquisizione digitale delle immagini fotografiche" emanata dall'Istituto centrale per il catalogo e la documentazione del Ministero dei beni e le attività culturali. Indica l'attività di eliminazione delle parti dell'immagine non contenenti informazioni essenziali.

apportati cambiamenti significativi, fatta eccezione per la creazione di nuovi indici di ricerca. Si è proceduto successivamente ad una personalizzazione dell'interfaccia, in quanto quella standard non si adattava ai requisiti del progetto. L'accesso alle immagini dei documenti è sempre mediato dalle descrizioni dei fascicoli. Questa fase ha richiesto più tempo, si è articolata lungo tutto l'arco del 2004 e si concluderà presumibilmente nel febbraio 2005.

- 5) L'ultima fase, non ancora conclusa, comprende il compimento del back-up di tutti i dati e l'ideazione di strategie di preservazione dei digital objects.

I capitoli in cui è strutturata la presente tesi riflettono a grandi linee le fasi in cui si è articolato il progetto.

Le traduzioni dall'inglese sono state fatte da me .

Le pagine web sono state controllate l'ultima volta il 16 gennaio 2005.

Ringrazio vivamente il direttore della biblioteca di Milano Bicocca, dott. Maurizio Di Girolamo, il prof. Natale Stucchi ed il prof. Mauro Antonelli per avermi proposto di partecipare al progetto, la dott. Barbara Marino e la dott. Verena Zudini per la loro preziosa collaborazione e l'amichevole rapporto instauratosi fra noi. Un sincero ringraziamento va inoltre ai colleghi Mariagrazia Pistelli e Luigi Zeliani per il loro valido contributo sugli aspetti tecnici di Aleph, a Bruno Pupo, ad Ivan Sciascia ed a Luigi Roli per la loro disponibilità e la pronta risposta a tutte le mie richieste di aiuto in campo informatico ed a tutti i colleghi della biblioteca per la simpatia che mi hanno sempre dimostrato.

CAPITOLO 1: L'archivio personale di Vittorio Benussi

CHI ERA BENUSSI

Vittorio Benussi era unanimemente considerato agli inizi di questo secolo uno dei più rigorosi e geniali psicologi sperimentali del suo tempo. Ben presto, tuttavia, la sua figura e la sua opera caddero nell'oblio, così da risultare oggi largamente sconosciute. Le storie della psicologia si limitano per lo più a ricordare Benussi come uno dei membri della scuola psicologica di Graz e come uno dei più importanti precursori del movimento gestaltista berlinese. Occasionalmente vengono altresì menzionate le sue ricerche sui «sintomi respiratori della menzogna», che posero, a livello applicativo, i fondamenti per lo sviluppo delle future “macchine della verità”¹.

Fritz Heider, ultimo allievo di Benussi all'Università di Graz, lo descriveva così:



He was an elegant-looking, lean person with a finely chiseled and melancholy face and a dry skeptical smile. He went around in a black laboratory smock and when he took a walk he put on a black hat with a wide brim and puffed on a long black cigar. One year there was a student who was often seen walking with him, much shorter, but in exactly the same black outfit. Benussi mostly worked in a darkened room where he had a cot, along with his apparatus, and he often spent the night as well the day there. He did not give very courses, perhaps because his health was not good. I remember one course in which he used

his students as subjects for a whole semester in an experiment on guessing the number of dots in a long series of patterns, and he did this without giving us any idea of the purpose of the experiment. I finally rebelled and told him humbly that I would like to learn psychology from him. He was

¹ VITTORIO BENUSSI, *Psychologische Schriften*, Amsterdam – New York, Rodopi, 2002 nella traduzione fornita dal curatore Mauro Antonelli.

very friendly, gave me a key to the laboratory, and said that all the apparatus was at my disposal but that he did not have much time.²

Nato a Trieste da Bernardo Benussi e Maria Rizzi nel 1878, trascorre l'infanzia e la giovinezza nella sua città natale. Il padre di Benussi, Bernardo (1846-1928), istriano, occupava con la sua attività di storico, condotta parallelamente alla professione di insegnante ginnasiale, un posto significativo nel mondo culturale giuliano. Proprio il padre costituisce un grosso punto di riferimento per Vittorio nel suo metodo di studio e di ricerca scientifica rigorosa. Nel 1896 Vittorio si trasferisce a Graz per studiare presso la Facoltà di Filosofia, qui entra in contatto con il professor Meinong con cui compie le prime ricerche sperimentali. Ben presto ottiene una delle borse di studio assegnate agli studenti più meritevoli. Dopo che egli ha conseguito la laurea ed il dottorato, Meinong gli affida la funzione di secondo assistente presso il Laboratorio di psicologia, accanto a Witasek. Le sue condizioni economiche lo costringono, però, ad accettare anche l'incarico come bibliotecario presso l'imperial- regia biblioteca dell'Università, impiego che manterrà fino al suo trasferimento a Padova.

Il lavoro diventa ben presto qualcosa di totalizzante, tale da assorbire completamente le sue energie: le uniche ore che passava lontano dal laboratorio erano quelle in cui lavorava in biblioteca. Tale intensa attività si rivela estremamente produttiva; grazie ad essa il suo nome comincia lentamente ad imporsi, anche a livello internazionale.³

A Graz ottiene nel 1905 la libera docenza e tiene i corsi di psicologia sperimentale fino al 1918. Durante questo periodo scrive opere che gli valgono la fama di psicologo singolare e di sperimentalista genialissimo. Come spiega lo storico della psicologia Mauro Antonelli:

Benussi incarna da un lato la mentalità scientifica rigorosa dello sperimentalista, attento ai dati effettivamente acquisiti o acquisibili con procedure metodologicamente corrette e sempre pronto a sottolineare i pericoli di teorizzazioni e generalizzazioni non supportate pienamente dal riscontro fattuale. D'altra parte è presente in Benussi una sensibilità artistica estremamente raffinata, riconoscibile nel suo entusiasmo per la

² FRITZ HEIDER, "Gestalt Theory: Early History and Reminiscences", in HENLE, JAYNES, SULLIVAN (eds.), *Historical Conceptions of Psychology*, New York, Springer, 1973, p. 66

³ VITTORIO BENUSSI, MAURO ANTONELLI (Hrsg), *op. cit.*

pittura, che praticava occasionalmente, e per la musica, di cui fu appassionato cultore⁴.

Di questa sua vena artistica si trovano tante prove tra i documenti dell'archivio. Nel 1914 compare il primo di una serie di lavori dedicato alla «psicologia forense». In esso Benussi espone i risultati delle sue indagini sui «sintomi respiratori» della consapevolezza di mentire e di dire la verità da parte di chi è chiamato a fornire una testimonianza di fronte ad un vasto pubblico. Tale interesse per la psicologia giudiziaria rinvia ai contatti di Benussi con Hans G. Groß, il fondatore della moderna criminalistica, dal 1905 ordinario presso l'Università di Graz. Groß fonda nel 1913 l'Istituto di criminalistica dell'Università di Graz, concepito come centro interdisciplinare di ricerca, e affida a Benussi il settore della psicologia applicata.

Gli interessi di ricerca di Benussi si aprono ora a settori nuovi ed inesplorati, posti ai margini della scienza psicologica ufficiale: è in questi anni che Benussi matura i primi interessi per i fenomeni ipnosuggestivi e manifesta le prime aperture verso la psicoanalisi freudiana, fatta oggetto di analisi approfondita durante il periodo di Padova.

Gli vengono offerte le cattedre di Praga e Vienna, ma rifiuta di prestare quella dichiarazione di lealismo al governo austriaco a cui era subordinata la nomina. Nel '19, dopo l'annessione di Trieste, si trasferisce in Italia dove, grazie all'aiuto dell'amico Sante De Sanctis, ottiene l'incarico come professore associato di Psicologia Sperimentale all'Università di Padova e lo mantiene fino al 1922, anno in cui gli viene conferita con una procedura straordinaria, per meriti eccezionali, la cattedra. Nella sua dimora di Padova fonda e dirige il Laboratorio di Psicologia Sperimentale che in soli otto anni, per la sua genialità tecnica e l'appassionata cura, diviene il migliore in Italia. Il 24 novembre 1927, a soli quarantanove anni, Vittorio Benussi viene trovato morto nel suo studio da alcuni suoi collaboratori: si è suicidato con una tazza di the e cianuro.

Quella di Benussi è senz'altro una personalità complessa ed affascinante, stravagante e a tratti anche contraddittoria, segnata da una sorte crudele. È soprattutto una personalità difficilmente collocabile o catalogabile.

⁴ *ibidem*

Trasferitosi appena diciottenne a Graz, Benussi vive fino in fondo il suo destino di «marginal man who stood between two cultures»⁵. Nonostante abbia vissuto più di vent'anni a Graz egli non riesce mai ad integrarsi completamente nel mondo austro-tedesco; quando poi, all'indomani del primo conflitto mondiale, Benussi si trova a svolgere la propria attività scientifica in Italia, egli vive con profondo disagio l'estraneità ad un ambiente culturale ed accademico che non era quello in cui era cresciuto ed in cui si era formato.

La sua persona appare sospesa non solo tra due culture, ma anche tra due epoche. Vittorio Benussi vive drammaticamente i conflitti e le contraddizioni del suo tempo, segnato dal progressivo dissolversi degli ideali e dei modelli positivisti. Egli incarna in misura esemplare la figura dell'«uomo senza qualità», educato al rigore delle discipline scientifiche esatte, ma al tempo stesso tormentato, inquieto, incapace di esorcizzare l'angoscia che deriva da ciò che sfugge al controllo della scienza e della ragione.⁶

NOTIZIE SULL'ARCHIVIO

L'unicità e la particolarità del materiale contenuto nell'archivio ha portato alla decisione di proteggerlo dai danni del tempo e, nel contempo, di valorizzarlo. La peculiarità deriva dall'originalità del suo produttore, dalla sua stravaganza, dalla sua passione per una disciplina ancora agli albori e dalla sua genialità in campo sperimentale, rimasta marginale a causa della situazione storica in cui si è sviluppata. L'unicità del materiale è un fatto oggettivo: soltanto un'esigua parte del materiale è stata pubblicata e questa dà solo in parte idea di quella che poteva essere la complessità di questo personaggio.

L'archivio personale di Vittorio Benussi accompagna buona parte della vita del suo produttore. Anche se sono presenti alcuni documenti con data precedente, si può far partire la datazione dell'archivio dal 1905, anno in cui il Nostro inizia a

⁵ FRITZ HEIDER, *op. cit.*

⁶ VITTORIO BENUSSI, MAURO ANTONELLI (Hrsg), *op. cit.*

lavorare all'università di Graz, fino alla morte avvenuta nel 1927: viene quindi coperto un arco temporale di quasi 25 anni.

Alla sua morte gli eredi non si interessarono del suo archivio, ne colse l'eredità intellettuale Cesare Musatti, prima suo allievo e poi stretto collaboratore, che conservò l'archivio del Maestro a Padova e poi lo portò con sé nei suoi vari spostamenti. Negli ultimi anni della carriera accademica Musatti lavorò all'Università degli studi di Milano e, quando andò in pensione, lasciò all'Istituto di Psicologia la sua biblioteca, parte del suo archivio personale e quello di Benussi. Allorché nel 1997 l'Istituto di Psicologia si smembrò, gran parte del patrimonio fu trasferito al nuovo Dipartimento di Psicologia dell'Università degli Studi di Milano Bicocca. Nel 2001 il Dipartimento di Psicologia decise di depositare il fondo presso la Biblioteca di Ateneo, ritenendo che quest'ultima fosse in grado di conservarlo in modo più adeguato.

Il fondo, di piccole dimensioni (ha una consistenza di 13 faldoni, pari a 1,95 metri lineari circa), è costituito in buona parte da materiale riguardante l'attività scientifica di Benussi, dalle sue pubblicazioni, dalla documentazione scritta per i corsi tenuti all'Università, dai protocolli e dai dati sperimentali che raccoglieva; vi si trovano inoltre il suo carteggio, documenti di varia natura e disegni sia suoi che dell'amico Gino Parin⁷.

L'archivio è stato suddiviso in 4 serie: Didattica, Carteggio, Pubblicazioni e Materiale sperimentale. La serie della Didattica è a sua volta suddivisa in 10 faldoni, mentre le altre tre serie sono composte da un unico faldone. In base alla lingua di composizione l'archivio può essere ulteriormente ripartito: fino al 1918 Benussi scrive quasi esclusivamente in tedesco, mentre dal 1919, quando

⁷ Gino Parin (Trieste 1876 - Bergen Belsen 1944) fu un pittore triestino, grande amico di Benussi. Dopo gli studi a Trieste, completò la sua formazione a Monaco di Baviera dove espose le sue opere con successo ottenendo una medaglia d'oro nel 1913. Fece delle mostre a Vienna ed a Trieste ed espose alla Biennale di Venezia tra le due guerre. Fu un buon ritrattista, accurato nella soluzione tecnica, ma senza forzature, e attento all'espressione dei soggetti che valorizzò con un buon senso della luce e con inquadrature originali. Operò inizialmente con stili classicheggianti e successivamente elaborò un verismo preciso e ricco di colore dai toni quasi musicali e fuso in eccellenti impasti per atmosfere ricche di intensa suggestione poetica. Con l'amico Vittorio condivise una tragica fine: ebreo fu perseguitato per motivi razziali, come altri triestini, e morì tragicamente in un campo di sterminio in Germania.

si trasferisce in Italia, usa quasi sempre la lingua italiana. È presente anche qualche documento in inglese e qualcuno in francese.

All'interno dei faldoni i documenti sono oggi suddivisi in fascicoli, ognuno contenuto in una camicia, in cui furono inseriti durante il primo riordino dell'archivio avvenuto nel 1982. Originariamente i documenti erano stati conservati in buste o in cartellette colorate, che in alcuni casi sono state mantenute per la loro particolarità.

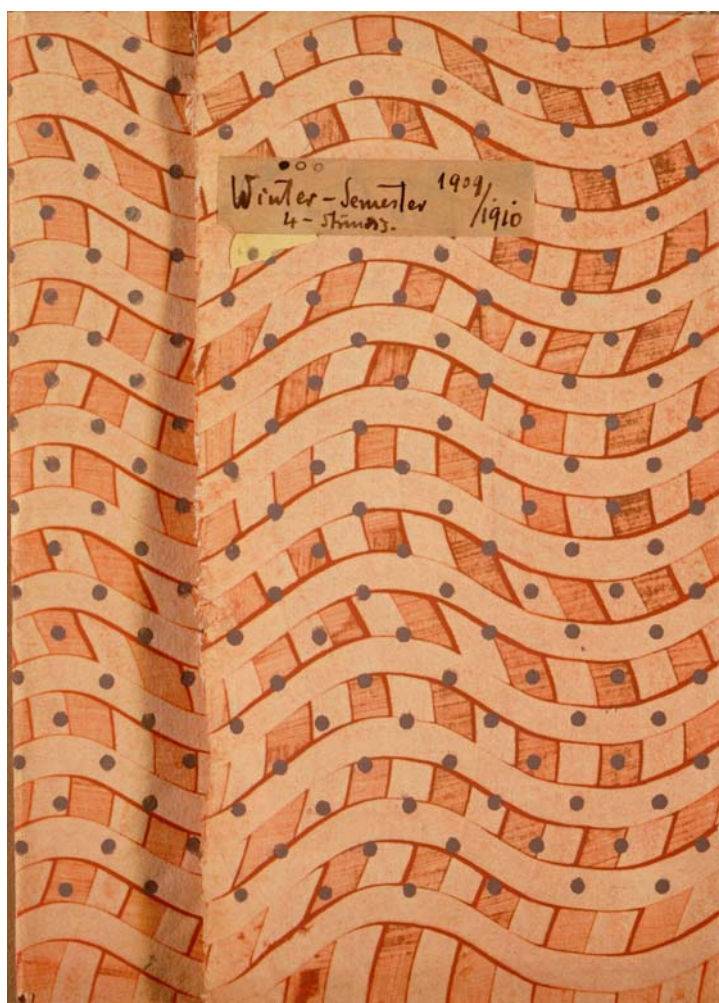


Fig. 1: Una delle cartellette usate da Benussi per conservare i suoi documenti.

Si è deciso di mantenere l'ordinamento già assegnato da Benussi, nonostante non rispecchi sempre l'ordine cronologico. Egli aveva infatti l'abitudine di riprendere argomenti già affrontati e di riunire quindi tutto il materiale inerente in un unico fascicolo.

I documenti presentano varie tipologie di supporto, nella maggior parte dei casi si tratta di fogli sciolti talvolta piegati in due a formare dei piccoli libretti; ma sono presenti anche alcuni quaderni ad anelli e block notes. Durante il periodo tedesco è frequente l'uso di carta velina, mentre nel periodo italiano lo psicologo scrive spesso su fogli rigati di carta di maggiore spessore e fa molto uso di piccoli cartoncini neri.

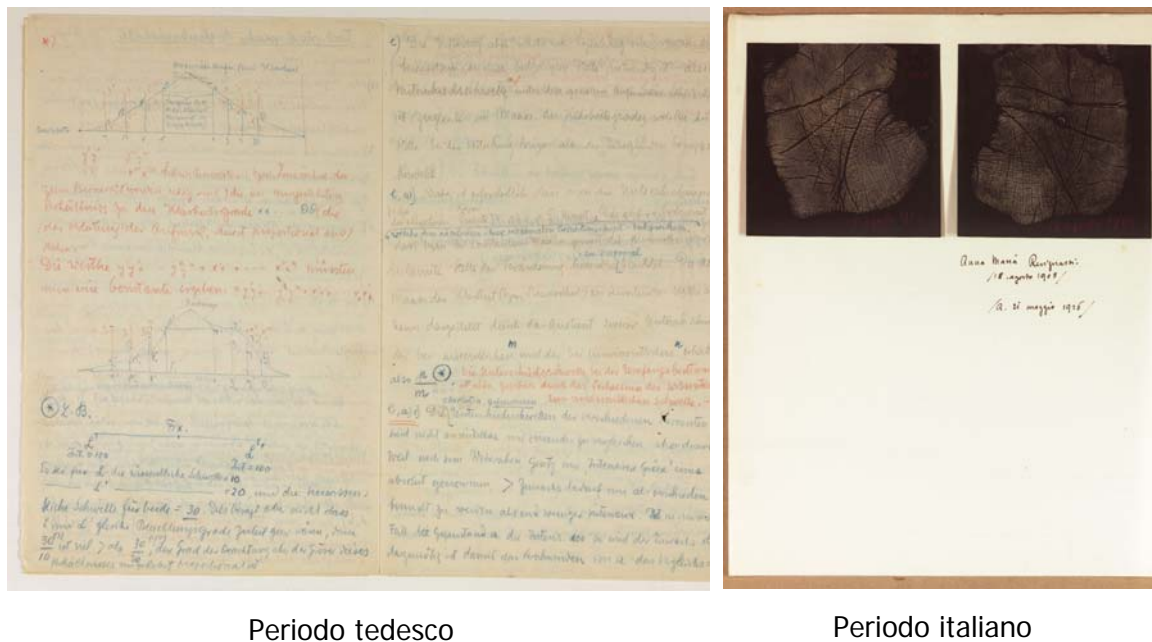


Fig.2: Esempio di due documenti

La maggior parte dei documenti è in forma manoscritta; ma, soprattutto nella serie delle Pubblicazioni, ci sono anche dattiloscritti e bozze di stampa.

Una differenza di comportamento tra il periodo tedesco e quello italiano si nota anche negli strumenti di scrittura; nel primo periodo è prediletto l'uso della matita, in molti casi colorata, mentre nel secondo è usato maggiormente l'inchiostro con il pennino.

ANALISI E SELEZIONE DI UN ARCHIVIO

La selezione rappresenta un aspetto critico nella definizione della fattibilità di un progetto in quanto permette di identificare i documenti o le raccolte valide per la digitalizzazione, di valutare il potenziale delle raccolte in rapporto ai bisogni

degli utenti e di decidere se pubblicare l'intera raccolta o se escludere alcuni materiali⁸.

È importante individuare i dati veramente essenziali e garantire un sistema di preservazione di questi dati con le nuove tecnologie, come sostiene Pierluigi Ridolfi quando scrive "È un errore pensare alla tecnologia come un mezzo per registrare tutto e per sempre". Questa affermazione si ricollega ad un'altra fatta nello stesso intervento:

Secondo alcune ricerche anche la memoria di ciascuno di noi [...] ha paradossalmente il compito primario di selezionare le informazioni, cancellandone la maggior parte per conservarne solo alcune, secondo personali, e per molti versi incogniti, criteri di valore. [...] Le informazioni entrano in una sorta di griglia selettiva a più livelli dove i processi di rinfresco riguardano quelli superiori, lasciando quelli inferiori ad una sorta di rumore di fondo⁹.

Affinché risulti efficace, la selezione deve coinvolgere tutti coloro che hanno una conoscenza dettagliata ed esperta del contenuto della raccolta e dei materiali, come il curatore e l'archivista.

I criteri di selezione dipendono dalle finalità del progetto, come anche dalle circostanze tecniche e finanziarie, dal copyright e dall'attività di altri progetti nello stesso ambito.

Il Minerva working group¹⁰ propone all'interno del suo *Handbook of good practice*¹¹ che almeno i seguenti criteri vengano presi in considerazione:

- accesso a materiali che altrimenti non sarebbero disponibili o sarebbero di limitata possibilità d'uso,

⁸ MICHELE SANTORO, "Digitalizzazione dei documenti e conservazione del digitale", corso Ebsco Training Center, Milano, 13-14 novembre 2003.

⁹ PIERLUIGI RIDOLFI, "Tecnologia e memoria: l'evoluzione dei calcolatori e la conservazione dell'informazione" in TULLIO GREGORY, MARCELLO MORELLI, *op. cit.*, p. 252.

¹⁰ Il Minerva Working Group è una rete costituita dai Ministeri degli stati membri dell'Unione Europea per discutere, correlare e armonizzare le attività condotte nella digitalizzazione dei contenuti culturali e scientifici per la creazione di una piattaforma comune e condivisa a livello europeo, raccomandazioni e linee guida sulla digitalizzazione, sui metadati, sull'accessibilità a lungo termine e sulla preservazione. Mira a coordinare programmi nazionali ed il suo approccio è fortemente basato sul principio dell'inserimento nelle attività nazionali di digitalizzazione. La rete messa in piedi da Minerva coinvolge Austria, Belgio, Danimarca, Finlandia, Francia, Germania, Grecia, Irlanda, Italia, Lussemburgo, Paesi Bassi, Portogallo, Spagna, Svezia e Gran Bretagna. Ogni working group è coordinato da uno dei diversi partner. Tra gli Stati Membri ci sono specifici sottogruppi di lavoro su uno stesso tema, che fa riferimento ad un coordinatore.

¹¹ MINERVA WORKING GROUP, "Good Practice Handbook"

<http://www.minervaeurope.org/structure/workinggroups/goodpract/document/goodpractices1_3.pdf>.

- accesso più ampio e più facile a materiale d'interesse,
- condizione degli originali,
- conservazione di originali delicati, rendendo disponibili versioni alternative,
- tema del progetto,
- copyright e diritto d'autore,
- disponibilità di versioni digitali esistenti,
- costi di digitalizzazione,
- presenza di materiale non considerato opportuno per l'accesso.

Paula De Stefano, nel suo contributo "Selection for Digital Conversion"¹², propone una prima superficiale selezione basata sul contenuto o sul formato, mentre riguardo agli altri criteri di selezione si muove nella stessa direzione delle linee guida proposte dal progetto Minerva.

L'accesso facilitato a materiali di ricerca remoti è una delle caratteristiche più allettanti delle risorse elettroniche. Un migliore accesso ad una collezione è un valido contributo per una comunità di utenti che può soddisfare un interesse sia a livello nazionale che internazionale. Steve Hensen¹³ sottolinea l'importanza di rendere accessibile in remoto materiale molto raro o unico contenuto in collezioni speciali, facilitando il lavoro di storici e ricercatori. Tuttavia non tutte le risorse meritano di essere sottoposte ad un lavoro così costoso ed impegnativo. Il contenuto dovrebbe essere, secondo lui, il principale fattore nella decisione di digitalizzare una collezione, specialmente se le attuali tendenze di ricerca e richiesta collaborano per renderlo un argomento di investigazione. Naturalmente il valore intellettuale da solo non è una ragione sufficiente per decidere di digitalizzare una raccolta, ma riveste una grande importanza; per questo la decisione di convertire al digitale deve essere approvata da specialisti del settore, che conoscano bene la collezione ed il suo valore scientifico e che ne valutino l'effettivo interesse presso la comunità scientifica. L'accesso elettronico ad una collezione si realizza in modo migliore quando viene sviluppato con la collaborazione di diverse istituzioni. Perché una

¹² PAULA DE STEFANO "Selection for Digital Conversion" in ANNE KENNEY, OYA RIEGER, *op. Cit.*, p. 11 e seguenti.

¹³ STEVEN HENSEN, *Primary Sources, Research, and the Internet: The Digital Scriptorium at Duke* <<http://www.scriptorium.lib.duke.edu/SCRIPTORIUM/hensen-aha97.html>>

biblioteca digitale sia fattibile e utile, una massa critica o un corpus di materiali di ricerca deve essere raccolto e reso accessibile a studiosi, facoltà, studenti e ad un'utenza generica. A tale proposito Paula De Stefano scrive:

La prospettiva di una massa critica di risorse digitalizzate è precisamente quello che più suscita interesse negli studiosi quando iniziano ad usare internet per il loro lavoro. Senza una massa critica non si può pienamente realizzare nessuna agevolazione in termini di tempo o convenienza relativa alla ricerca basata sul web. Quando i ricercatori e le facoltà parlano in termini di contenuto, l'accuratezza sembra la chiave per soddisfare i loro bisogni nella ricerca elettronica.¹⁴

Un'altra tradizionale metodologia di selezione si basa infatti sul contributo di studiosi ed altri ricercatori. Nel contesto di un progetto allargato ad un'intera comunità scientifica, il progetto Perseo¹⁵ alla Tufts University impiega attivamente un consiglio di esperti del settore per sviluppare il contenuto delle collezioni digitali, soddisfacendo in questo modo i bisogni formativi.

Infine si devono considerare le condizioni dei supporti. Il continuo uso del materiale archivistico e librario contribuisce alla sua scomparsa. Se il materiale è fragile il maneggiamento può essere causa dell'irrevocabile perdita dello stesso, motivo per cui la digitalizzazione serve a creare una copia surrogato utilizzabile in sostituzione dell'originale.

LA SELEZIONE NELL'ARCHIVIO BENUSSI

L'archivio di Benussi è stato acquisito quasi interamente in nome della sua unicità e dell'interesse che potrà suscitare nella comunità degli storici della psicologia e degli psicologi in generale. Si è deciso di non procedere alla digitalizzazione soltanto della serie contenente il materiale sperimentale non adatto alla visione statica e bidimensionale.

Una scelta di questo tipo è stata resa attuabile grazie ad una fortunata unione di fattori che ne permettevano la completa acquisizione in formato digitale:

- un'adeguata copertura finanziaria. I fondi del cofinanziamento ministeriale, grazie anche ad un contratto vantaggioso con la ditta che

¹⁴ PAULA DE STEFANO, *op. cit.*, p. 16.

¹⁵ Progetto Perseo <<http://www.perseus.tufts.edu>>

ha effettuato la digitalizzazione, consentivano di portare avanti questa scelta.

- le condizioni del materiale. In generale i supporti si presentavano in condizioni abbastanza buone, assicuravano quindi sul fatto che il processo di digitalizzazione non avrebbe arrecato loro danno.
- la mancanza di copyright. Benussi è morto da 77 anni, dunque i suoi eredi e gli editori presso cui stampò le sue pubblicazioni non possono più vantare diritti di tipo economico.
- l'assenza di altre risorse digitali nello stesso ambito e sugli stessi argomenti.

LA CONSERVAZIONE DELL'ARCHIVIO CARTACEO

Un esame sulla condizione degli originali è essenziale in quanto influenza le tecniche di scansione e determina la possibilità di ottenere una buona immagine digitale.¹⁶

Prima di avviare un progetto di digitalizzazione si deve valutare lo stato fisico del materiale: la qualità delle pagine e delle carte, l'eventuale presenza di pagine sporche o ingiallite e di originali degradati o deteriorati.

Nella conservazione del materiale cartaceo si deve prestare attenzione al deterioramento della carta ed alle interazioni tra questa e l'inchiostro che si può scolorire o può danneggiare il suo supporto.

La carta subisce processi chimici e fisici, determinati dalla sua composizione e dalle reazioni con l'ossigeno, l'umidità e le sostanze inquinanti dell'aria, che se non possono essere fermati, possono quantomeno essere rallentati.

Per ritardare il processo di deterioramento è necessario che il luogo di deposito soddisfi determinate condizioni: ambienti muniti di filtri contro l'inquinamento dell'aria, una temperatura costante al di sotto dei 18° C ed un tasso di umidità relativa intorno al 30%.

CONDIZIONI DELL'ARCHIVIO

¹⁶ MICHELE SANTORO, "Digitalizzazione dei documenti e conservazione del digitale", corso Ebsco Training Center, Milano, 13-14 novembre 2003.

Vittorio Benussi ha costituito il suo archivio a margine di un periodo critico per la fabbricazione della carta. Per fortuna lo psicologo sceglieva carta di buona qualità che al momento non ha subito un vistoso deterioramento. Sono molto poche le carte che hanno assunto un colore giallo intenso e che rischiano di sbriciolarsi tra le mani; si tratta in particolare di ritagli di giornale. Per il resto la carta ha mantenuto la sua resistenza; verranno comunque prese delle misure per cercare di rallentare il processo di invecchiamento. Alla fine del progetto il fondo verrà collocato nel deposito della biblioteca all'interno di armadi di metallo. Il deposito è un ambiente climatizzato, dove la temperatura viene tenuta costante a circa 18° C, è deumidificato ed è fornito di doppie porte tagliafuoco. All'interno del deposito è stata creata un'area per la consultazione del materiale, quindi l'archivio non subirà – si spera - ulteriori spostamenti.

CAPITOLO 2: La digitalizzazione

Lo schermo di un computer è costituito da una serie di piccole "celle", ciascuna in grado di mostrare il nero, il bianco, il grigio o il colore. Uno schermo di buon livello ha 1024 celle che lo attraversano orizzontalmente e 768 che lo attraversano verticalmente: questo significa che lo schermo ha una risoluzione di 1024x768. Queste celle prendono il nome di picture elements, di solito abbreviato in pixel. Ciascun colore o gamma di colori è dato da un codice del computer memorizzato come codice binario, cioè come una serie di 0 e di 1. Al livello più semplice, questo codice è dato da un codice ad una cifra, ad esempio lo 0 corrisponde al nero, mentre l'1 corrisponde al bianco. Se al computer viene detto che un certo pixel ha il codice 0 associato ad esso, il computer riempie quel pixel con il nero e così via. L'immagine che ne risulta è dunque un'immagine ad un bit (per ciascun pixel è necessario un solo bit) e si dice che ha una "profondità" di due: ci sono cioè due possibili valori con cui esso può essere riempito o a cui può essere associato, 0 e 1. I codici dei pixel sono memorizzati ed elaborati nella memoria del computer come una griglia o una mappa di bit (bitmap); se viene istruito dal software, il computer converte questi codici negli appropriati colori e li proietta sullo schermo, cosicché da una certa distanza possano sembrare un'immagine normale (questo processo è realizzato da un pennello elettronico che riempie ciascun pixel orizzontalmente da sinistra verso destra e verticalmente dall'alto verso il basso).

Solitamente l'immagine digitale è la rappresentazione di un originale analogico, ma in realtà l'immagine presentata dal monitor di un computer o dalla stampa non riflette le qualità dell'originale, è "il risultato cumulativo della qualità dell'immagine surrogato, della risoluzione della scansione, della profondità dei pixel, del tipo di scanner, dell'abilità dell'operatore, dell'uso di descrittori del colore indipendenti dal dispositivo e dal colore"¹.

Il processo attraverso il quale si giunge ad una rappresentazione digitale da un originale analogico viene chiamato digitalizzazione.

¹ Giornata di studio sulla Digitalizzazione e Archiviazione delle immagini, ITIM, Milano, 23 marzo 2000 <<http://www.itim.mi.cnr.it>>.

IN CHE COSA CONSISTE LA DIGITALIZZAZIONE

Secondo quanto già accennato in precedenza il processo di digitalizzazione consiste nell'estrazione di informazioni da un'immagine originale e nella loro conversione in forma binaria tramite una periferica di input, come uno scanner o una macchina fotografica.

La differenza tra analogico e digitale corrisponde alla differenza fra una rappresentazione continua e una rappresentazione discreta di determinate grandezze.²

La conversione del segnale è la misurazione secondo una frequenza fissa, cioè secondo un'unità di tempo costante, dell'ampiezza del segnale stesso. Questo primo passo, chiamato campionamento (sampling), rappresenta una prima selezione dell'informazione contenuta nel segnale analogico. Minore è l'intervallo tra una misurazione e l'altra, maggiore sarà la precisione della rappresentazione digitale. Il secondo passo rappresenta anche il secondo livello di approssimazione nella conversione: è necessario attribuire un valore discreto (quantization), che non potrà mai essere esattamente coincidente con la realtà. L'ultimo passaggio del processo di digitalizzazione consisterà nell'espressione dei dati ottenuti dalle misurazioni in una serie numerica (code word) leggibile dalla macchina.³

SCELTA DEI PARAMETRI

Nella pianificazione è importante stabilire quali sono gli scopi iniziali e quali quelli a lungo termine, identificare i bisogni istituzionali, decidere se le immagini avranno un uso specifico oppure entreranno a far parte di un archivio generico.

Pochi standard regolano la creazione e l'uso delle immagini digitali e finora non è stato elaborato un approccio uniforme che si adatti a tutte le circostanze. Nel decidere come digitalizzare e presentare il materiale bisogna tenere conto di una serie di caratteristiche, come la natura dei documenti, gli obiettivi e le risorse, le

² FABIO CIOTTI, GINO RONCAGLIA, *Il mondo digitale. Introduzione ai nuovi media*, Roma – Bari, Laterza, 2000.

³ ALBERTO SALARELLI, ANNAMARIA TAMMARO, *op. cit.*, p. 58.

capacità tecnologiche e le finalità del progetto.

Il primo passo sarà quindi quello di definire parametri di qualità, leggibilità, velocità nella reperibilità e fedeltà all'immagine.

Il JISC, il Joint Information Systems Committee, consiglia innanzitutto:

Where possible proprietary standards will be avoided and de facto standards, or those propagated by standards organisation, will be preferred.⁴

Attualmente esistono diverse teorie su quali devono essere i parametri, cresce però l'idea che sia necessario creare delle immagini digitali il più possibile ricche di informazioni in modo da prolungarne l'uso nel tempo. I file digitali hanno lo scopo di sostituire o ridurre il ricorso all'originale, a patto che i loro surrogati offrano una rappresentazione accurata e veritiera. La qualità, l'utilizzo, la visione e la gestione sono strettamente dipendenti dalla qualità della scansione iniziale, come si dirà successivamente in modo più approfondito.

Nella scelta dei parametri per la conversione digitale alcuni fattori sono determinanti nelle decisioni riguardanti l'acquisizione e la presentazione delle informazioni⁵:

- La conservazione, cioè la tutela degli originali, deve condizionare le scelte riguardanti la digitalizzazione.
- L'aspetto economico: una maggiore risoluzione, profondità dei bit, o la scelta della scansione a colori piuttosto che in bianco e nero rallentano la produzione e alzano i costi di scansione e conservazione. Bisogna tuttavia tenere conto che una scansione di migliore qualità innalza i costi inizialmente, ma permette di recuperare tali costi in termini di longevità e facilità di rielaborazione delle immagini.
- La tecnologia: è inutile basare i parametri di conversione sulle

⁴ TECHNICAL ADVISORY SERVICE FOR IMAGES, "Why "Archive Standards" " in *Technical Advisory Service for image , Advice Paper* <<http://www.tasi.ac.uk/advice/creating/pdf/why-archive.pdf>>. Il Tasi, Technical Advisory Service for Images, è un servizio, finanziato dal JISC, che fornisce consigli e indicazioni su questioni come la creazione di immagini digitali, il reperimento delle immagini per gli utenti, l'uso di immagini digitali come supporto all'insegnamento, all'apprendimento ed alla ricerca, la gestione di progetti di digitalizzazione su piccola e larga scala. Gli obiettivi del servizio sono incoraggiare la creazione e l'uso di immagini di alta qualità in campo scientifico, promuovere l'uso di standard, incoraggiare i collegamenti e la costruzione di una comunità digitale.

⁵ ANNE KENNEY, OYA RIEGER, *op. cit.*

caratteristiche tecniche di monitor, stampanti e reti, perché si corre il rischio che le immagini non siano più utilizzabili nel momento in cui le caratteristiche tecniche di tali strumenti migliorano.

- I bisogni dell'utenza: la prospettiva e le necessità degli utenti devono ispirare i parametri della conversione e dell'accesso alle immagini. Per identificare tali bisogni si dovranno identificare gli utenti e gli usi potenziali del materiale.

Si passerà quindi a valutare il contenuto informativo e le caratteristiche dell'originale; bisogna cioè tenere conto della sua consistenza (numero delle immagini), della dimensione degli originali, della quantità di informazioni riguardanti i documenti, delle sue qualità fisiche (rappresentazione, tipo di fonte e condizioni del supporto), dell'ordinamento e della classificazione dei documenti, delle capacità visive dell'uomo (rappresentazione tonale e percezione di colore e dettagli).

La determinazione del tipo e del livello di informazione, presente nell'originale, di ciò che è essenziale al suo significato e che quindi deve essere contenuto nel file digitale, coinvolge valutazioni oggettive e soggettive. Una valutazione oggettiva può essere fatta sottoponendo tutte le caratteristiche delle immagini a dei test, mentre il contenuto informativo può essere determinato soggettivamente dalla capacità visiva umana. Sebbene ci siano stampanti che arrivano ad usare 1.200 dpi, l'occhio umano non riesce a percepire la differenza tra un'immagine digitale con una risoluzione di 600 dpi ed una a risoluzione 1.200 dpi⁶; di conseguenza digitalizzare ai massimi livelli diventa inutile per la dimensione dei file che ne risulterebbero, per i tempi più lunghi di scansione, per gli scopi e le destinazioni finali.

Le immagini ad alta risoluzione infatti danno vita a file molto grandi e di conseguenza costosi da memorizzare, inoltre le macchine digitali ad alta risoluzione spesso impiegano diversi minuti per catturare ciascuna immagine, con un notevole spreco di tempo; quindi, per esempio, se l'immagine sarà usata soltanto per il web

⁶ ANNE KENNEY, OYA RIEGER, *op. cit.*, p. 29.

o sarà stampata e poi eliminata o se proviene da un documento di bassa qualità non conviene digitalizzarla a 24 bit.

Una valutazione soggettiva diventa più problematica quando l'obiettivo è la leggibilità più che la fedeltà. Per il materiale testuale il grado di leggibilità dovrebbe essere definito come la variazione tra la qualità marginale e l'alta fedeltà; mentre è più difficoltoso dare un giudizio per l'informazione non testuale. La Library of Congress, per esempio, ha digitalizzato la sua collezione di musica a stampa con una risoluzione di 400 dpi per catturare dettagliatamente la notazione.⁷

Una standardizzazione ancora inferiore può essere applicata quando si prendono in considerazione altri requisiti, come il tono ed i colori in quanto questi dipendono strettamente dal contesto.

Il processo di assegnazione degli attributi dei documenti catturati agli equivalenti digitali comincia con un'attenta valutazione di quali aspetti dell'originale influenzeranno l'approccio della scansione e di quali di essi debbano essere rappresentati nel surrogato digitale. Il secondo passo è determinare se si possano rappresentare quegli attributi tramite misurazioni oggettive.

Il TASI offre delle linee guida, efficacemente riassunte in una tabella, su quelle che dovrebbero essere le operazioni da svolgere prima di partire con il processo di acquisizione delle immagini.

Pre Image Capture	Action
Equipment, Software and Environment	All equipment should be tested and calibrated. The capture studio should be prepared with consistent controlled lighting and the digitisation team trained in the use of all capture equipment and software within an established and standardised workflow.
Testing and Establishing Operational Specifications for image capture	Before any image capture can be undertaken it is imperative that the operators know the required File sizes and File Types. Once these have been established they will become part of the Operational Specifications. These should be established and agreed with all interested parties before the onset of project.
Establish capture workflow	All tasks within the capture workflow should be considered and a manual of 'Good Practice'

⁷ La documentazione riguardo al progetto American Memory della Library of Congress è reperibile all'indirizzo <<http://memory.loc.gov/ammem/about/techIn.html>>.

	established to guide all operators.
Consider the use of an 'Objective' or 'Subjective' colour management system within the capture workflow	<p><i>'Objective'</i> capture: In an objective capture workflow the colour of the digital images can be objectively measured and compared against the original work. This is normal within a 'Direct Digital Capture' workflow where there has not been any intermediary image. This is a typical capture workflow when working with a digital camera or when scanning direct from the original. An objective capture workflow should be calibrated and characterised using the ICC profiling system and then operator colour adjustment prevented.</p> <p><i>'Subjective'</i> capture: In a subjective capture workflow the colour of the digital image can only be compared against another 'analogue intermediary' image (often a copy transparency of the original work). As this analogue intermediary has already introduced a subjective element to the digital image, it will be necessary to allow the operator to make colour adjustments according to his skill and understanding of the original work to correct any possible fault within the 'analogue intermediary'. This 'subjective' workflow is typical of any system that is working with some form of copy of the original.</p>
Collect and prepare original images	Agree with curators / conservators that all originals are in a stable state and can withstand the necessary handling. If this is not the case, now is the time to have any required conservation work done. It is very important that originals can be provided for capture at a sufficient rate to prevent any bottleneck in workflow.
Clean originals	If the original is a 'work of art' then the method of cleaning will be dependent upon the requirements of the curator, conservator or owner. However any cleaning at this stage will provide higher quality and speed than having to digitally 'clean' them after capture.
Clean Capture Device	Scanners must be kept scrupulously clean, however cameras need to be cleaner still. All capture equipment should be externally cleaned every day and internally cleaned as required, which is normally at least each week. All cleaning should be done in strict accordance with the manufacturer's recommendations.

Tabella 1: Operazioni precedenti l'acquisizione delle immagini⁸

⁸ "Basic Guidelines for Image Capture and Optimisation" in "Technical Advisory Service for image , Advice Paper" <<http://www.tasi.ac.uk/advice/creating/pdf/img-capt.pdf>>.

IL LABORATORIO DI DIGITALIZZAZIONE

Prima di approntare un ambiente di lavoro bisogna valutare attentamente se è più appropriato affidare ad esterni il processo di acquisizione delle immagini piuttosto che digitalizzarle internamente. La scelta di affidare il lavoro di digitalizzazione ad una società di servizi è più conveniente nel caso di progetti aventi una quantità di documenti da digitalizzare piuttosto limitata, infatti le società di outsourcing sono costose e si deve creare un sistema di monitoraggio della qualità del loro lavoro. Inoltre il trasporto ed il maneggio da parte di terzi costituiscono fattori di rischio per il materiale.

Digitalizzare all'interno dell'istituzione è sicuramente più vantaggioso se si hanno già le competenze ed i mezzi adeguati. In caso contrario un approccio ibrido, in cui si divide il lavoro e si combinano i benefici delle risorse interne ed esterne, secondo la documentazione del TASI⁹, può risultare la scelta migliore.

Elemento non trascurabile è la scelta di un ambiente di lavoro adeguatamente illuminato, dove devono essere sistemati l'hardware per il processo di digitalizzazione (scanner, macchine digitali, postazioni di copia), il computer a cui collegare l'hardware, un software per la cattura delle immagini e per il fotoritocco, un masterizzatore per CD e DVD, un software di gestione dei metadati e uno per il controllo di qualità.

Area di lavoro

Lo spazio di lavoro deve essere abbastanza ampio da permettere agevolmente di maneggiare, ordinare e mettere al loro posto i materiali durante la digitalizzazione; la dimensione di tale spazio potrà variare in base al tipo di materiale che deve essere acquisito. Un dettaglio da non trascurare è anche la posizione e la disponibilità delle prese elettriche e telefoniche indispensabili per lavorare.

Illuminazione

Lavorare con le immagini digitali normalmente coinvolge anche la visione e la valutazione dei colori, pertanto le condizioni di illuminazione intorno all'area di

⁹ TECHNICAL ADVISORY SERVICE FOR IMAGES, "Setting up a Workspace for Digitization" in *Technical Advisory Service for image , Advice Paper*
<<http://www.tasi.ac.uk/advice/creating/pdf/workspace.pdf>>

lavoro devono rimanere costanti così che il colore percepito sia una fedele rappresentazione del colore visto.

Il TASI consiglia alcuni parametri per cercare di rendere la luce standard¹⁰:

- dipingere le pareti della stanza in cui si lavora di bianco o di grigio,
- eliminare tutte le luci estranee all'area di lavoro, chiudendo anche le finestre per evitare la luce del giorno,
- usare luci dai colori standard nell'area di lavoro (se la luce ha un colore standard un foglio bianco sul monitor ed un foglio di carta devono avere lo stesso colore),
- prevedere l'uso di filtri per evitare la distorsione del colore,
- assicurarsi che il livello della luce sia abbastanza scuro e che il contrasto del monitor sia ben visibile, ma nello stesso tempo abbastanza chiaro da vedere facilmente che cosa si sta facendo,
- fare in modo che le luci non si riflettano sui monitor dei computer dal punto di vista dell'operatore,
- scegliere come sfondo per il desktop del sistema operativo un grigio medio, in modo da avere uno sfondo ben bilanciato per le immagini.

Periferiche di acquisizione

Le periferiche di acquisizione si dividono in due categorie in base al modo in cui avviene il processo di campionamento. Lo scanner normalmente impiega un line-array sensor che cattura una linea di pixel alla volta, per cui il sensore si deve poter spostare da una parte all'altra del piano dell'immagine. La fotocamera digitale, invece, è fornita di un area-array sensor, che cattura i valori dei pixel all'interno dell'immagine con un'unica esposizione. Mentre la fotocamera ha un meccanismo di focus per regolare la distanza tra l'oggetto ed il sensore, lo scanner non consente di regolare questo parametro. Dalla scelta dello scanner o della camera dipende in buona parte il successo del progetto. Tutte le periferiche sono composte da un sistema ottico, da un sensore luminoso, da un'interfaccia e da un

¹⁰ TECHNICAL ADVISORY SERVICE FOR IMAGES, "Setting up a Workspace for Digitization" in *Technical Advisory Service for image , Advice Paper*
<<http://www.tasi.ac.uk/advice/creating/pdf/workspace.pdf>>

software. La qualità della scansione dipende innanzitutto dalla qualità del rilevatore ottico e luminoso, mentre la produttività è strettamente dipendente dall'interfaccia e dal software.

Nella valutazione e scelta dello scanner devono essere considerati alcuni elementi, tra cui fondamentale è la risoluzione che è la misura della capacità di catturare un dettaglio nell'originale. Si parla solitamente di risoluzione "ottica" e "interpolata": "[...]the optical resolution is limited by the number of elements (pixels) in the detector array and how the array is moved relative to the image"¹¹, mentre la risoluzione interpolata viene definita come "a result of the scanner software 'guessing' the values between pixels and presenting these intermediate values as 'real' values"¹². Un'altra considerazione fondamentale è che lo scanner dovrebbe essere largo almeno quanto il documento più grande che deve essere digitalizzato, in modo da evitare di catturare le immagini a mosaico, cioè un pezzettino alla volta, riunendole poi virtualmente in un'unica immagine. Naturalmente la scelta deve ricadere sull'hardware più potente e flessibile che possa essere usato.

Piano fotografico

Il piano fotografico e il piano del materiale da digitalizzare devono essere esattamente paralleli, altrimenti l'immagine risulterà distorta.

Software per il fotoritocco

È importante acquistare un adeguato software per il ritocco dei file master con la capacità di aprire file molto grandi, di modificare la risoluzione e la profondità del colore, di salvare molteplici versioni differenti, di selezionare e copiare una parte delle immagini e di salvarle come un'altra immagine, di esportare le immagini in differenti formati includendo gli standard web JPEG e GIF¹³.

Computer

Un computer con una ingente capacità di memorizzazione deve essere connesso alla strumentazione; inoltre deve essere fatto regolarmente un back-up dei dati.

¹¹ TECHNICAL ADVISORY SERVICE FOR IMAGES, "Scanners" in *Technical Advisory Service for image*, *Advice Paper* <<http://www.tasi.ac.uk/advice/creating/pdf/scanner.pdf>>

¹² *ibidem*

¹³ TECHNICAL ADVISORY SERVICE FOR IMAGES, "Setting up a Workspace for Digitization" in *Technical Advisory Service for image*, *Advice Paper* <<http://www.tasi.ac.uk/workspace.pdf>>

LA RISOLUZIONE

Elemento fondamentale per la qualità dell'immagine è la risoluzione. All'interno del processo di digitalizzazione devono essere presi in considerazione due tipi di risoluzione: la risoluzione spaziale (cioè il numero dei pixel) e la risoluzione del colore (cioè la profondità dei bit).

La risoluzione spaziale si riferisce alla frequenza con cui viene eseguito il campionamento, cioè al numero di pixel usati per convogliare l'immagine stessa. Viene espressa in dpi (dots per inch, punti per pollice), in cui il pollice si riferisce ad un pollice del documento originale, oppure in ppi (pixel per pollice) in riferimento alla risoluzione a schermo o dell'oggetto digitale. Nella scansione di un'immagine lunga 10 pollici e larga 5 pollici, per esempio, lo scanner o la macchina fotografica digitale possono essere programmati per digitalizzare a 100 punti per pollice (100 dpi), il che significa che ciascun pollice quadrato del documento originale può essere catturato a una risoluzione di 100 pixel per 100 pixel moltiplicando i punti per pollice per la grandezza del documento originale. Più punti per pollice vengono digitalizzati, più alta è la qualità dell'immagine in quanto vengono catturate più informazioni relative a ciascun pollice quadrato del documento originale.

Se si prendono per esempio due foto a stampa, una di 4" x 5" e un'altra di 8" x 10", entrambe fotografate in una scala di toni di grigio alla stessa risoluzione (300 dpi), la dimensione dei file per l'immagine 8" x 10" è quattro volte maggiore (7,2 Mb in confronto a 1,8 Mb). Se, invece, la dimensione dei file di ciascuna immagine è la stessa, la risoluzione dell'immagine 8" x 10" sarà la metà di quella dell'immagine 4" x 5" (150 dpi in confronto a 300 dpi). La formula per calcolare la dimensione dei file sarà la seguente:

Dimensione file = altezza x larghezza x profondità dei bit x dpi/8 bit per byte¹⁴

Il Research Libraries Groups (RLG) della Library Federation propone altre due

¹⁴ ANNE KENNEY, OYA RIEGER, *op. cit.*, p. 32.

formule per tale calcolo¹⁵:

Dimensione file = dimensione dei pixel x numero dei canali

o ancora:

D.F. = dimensione in pollici x risoluzione per pollice x numero dei canali

Molte istituzioni, per ovviare al problema delle grosse dimensioni dei file, hanno deciso di digitalizzare i documenti di grosso formato ad una risoluzione inferiore rispetto a quella scelta; questa decisione comporta però una diminuzione del dettaglio nell'immagine.

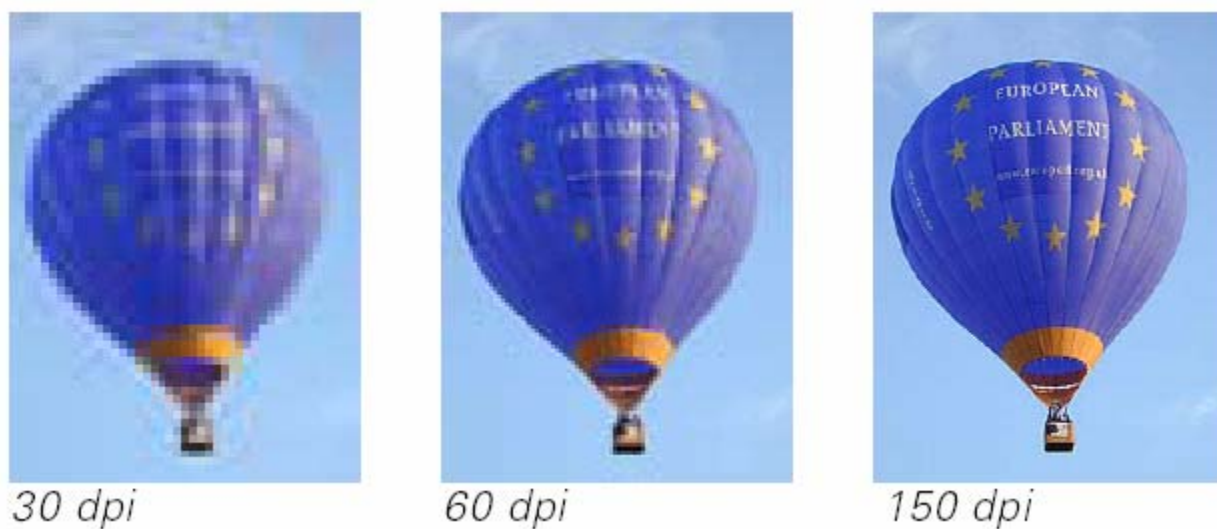


Immagine 1: l'aumento della risoluzione dà qualità più alta¹⁶

La cattura delle immagini deve sempre avvenire alla più alta risoluzione ragionevole¹⁷. Per questo il Minerva working group non fissa un valore, in quanto

¹⁵ DIGITAL LIBRARY FEDERATION – RESEARCH LIBRARIES GROUP, Guide to quality in Visual Resource Imaging, 2000. <<http://www.rlg.org>>.

¹⁶ TECHNICAL ADVISORY SERVICE FOR IMAGES, "The Digital Image" in *Technical Advisory Service for image*, Advice Paper <<http://www.tasi.ac.uk/advice/creating/pdf/image.pdf>>.

¹⁷ MINERVA WORKING GROUP, "Good Practice Handbook" <http://www.minervaeurope.org/structure/workinggroups/goodpract/document/goodpractices1_3.pdf>.

esso dipende dalla natura del materiale da digitalizzare e dagli usi che se ne faranno. La regola per stabilire la giusta risoluzione consiste nel valutare se la scansione ad una risoluzione più alta non offra maggiori informazioni rispetto ad una più bassa. Per individuare un adeguato grado di risoluzione alcune istituzioni hanno confrontato le stampe dell'immagine digitale con l'originale o con una sua riproduzione analogica. Solitamente la risoluzione si decide in base alle caratteristiche della collezione ed all'uso che si deve fare delle immagini, ma non sempre è possibile sapere in anticipo quale sarà l'output o quali saranno le dimensioni dell'immagine. Per questo motivo anche il TASI suggerisce:

Original works should be captured and archived in the highest colour and spatial resolution allowed by a project's budget. It is important that this standard is 'good enough' to provide the information needed to create all images required by the project both in the present as well as into the future. This archival image, or a copy of it, can then be optimised to create a surrogate image for any form of output.¹⁸

Questo consiglio deriva dal fatto che ogni periferica di output richiede un certo numero di informazioni per fornire la migliore qualità possibile.

Risoluzione per il web

I monitor delle ultime generazioni possono essere facilmente impostati per visualizzare un'ampia gamma di risoluzioni entro le loro potenzialità fisiche. Ogni pixel nella grafica del web corrisponderà ad un pixel sullo schermo senza che il browser sia ammaestrato a fare altrimenti dal codice html delle pagine visualizzate. Prendendo ad esempio un'immagine larga 200 pixel e lunga 100, visualizzata su un monitor di 15 pollici, quando la risoluzione del monitor sarà settata a 640 x 480 l'immagine sarà visualizzata con una dimensione di circa 3.5 pollici in larghezza e 1.75 pollici in altezza. Lo stesso monitor settato ad una risoluzione di 1280 x 1024 pixel visualizzerà la stessa immagine con una dimensione di circa 1.75 pollici in larghezza e 0.75 pollici in altezza. Quindi perché un'immagine appaia accettabile per una vasta gamma di settaggi dovrebbe essere creata in modo da assicurare che informazioni sufficienti siano comunicate ad ogni

¹⁸TECHNICAL ADVISORY SERVICE FOR IMAGES, "The Digital Image" in *Technical Advisory Service for image*, *Advice Paper* <<http://www.tasi.ac.uk/advice/creating/pdf/image.pdf>>.

utente.

Risoluzione per la stampante

Le stampanti a getto d'inchiostro usano una dimensione costante di punti molto piccoli e variano la frequenza dei punti per costruire il colore (questo processo è chiamato "modulazione della frequenza"). La modulazione permette stampe di migliore qualità da file più piccoli necessitando di una risoluzione non molto alta. Una risoluzione tra i 175 ed i 225 pixel per pollice forniranno le informazioni necessarie per ottenere una stampa di buona qualità.

COLORE vs B/N

L'occhio umano percepisce il colore quando la combinazione delle lunghezze d'onda della luce visibile colpisce la retina, ma molto spesso combinazioni di lunghezze d'onda diverse producono la medesima percezione del colore¹⁹. La percezione del colore può essere molto soggettiva in quanto dipende dalle caratteristiche biologiche e fisiologiche dell'osservatore, infatti ogni individuo recepisce i segnali in maniera differente.

Esistono tre tipi di recettori nell'occhio, di conseguenza la sensazione del colore può essere descritta attraverso tre valori. Ad ogni colore percepibile vengono assegnati tre codici o tre insiemi di coordinate. Ognuna di queste è un asse in uno spazio tridimensionale e la gamma di tutti i colori percepibili riempie questo spazio. L'insieme dei codici o delle coordinate cambia a seconda del color model, cioè dell'insieme delle proprietà dello spettro del colore in termini standard.

Esistono diversi color model, di cui RGB e CMYK sono i due principali. Il color model RGB, chiamato anche additive color system, è costituito da rosso, verde e blu. Questo modello combina la luce per produrre una gamma di colori e, mescolando i colori primari, crea i colori complementari. Scanner e monitor usano questo tipo di modello.

Il modello CMYK (composto da ciano, magenta, giallo e nero) parte invece dal principio che gli oggetti assorbono alcune lunghezze d'onda e ne riflettono altre.

¹⁹ DIGITAL LIBRARY FEDERATION – RESEARCH LIBRARIES GROUP, Guide to quality in Visual Resource Imaging, 2000. <<http://www.rlg.org>>

Esso usa i pigmenti colorati e coloranti che filtrano la luce portando via il colore dalla luce bianca. Su questo modello si basa in particolare la stampa professionale.

Il color gamut è invece la gamma totale dei colori riprodotti da una periferica.

Il colore può essere definito da tre proprietà: tonalità, saturazione e luminosità. La tonalità dipende dal modo in cui la lunghezza d'onda della luce colpisce un oggetto; la saturazione è l'intensità o la purezza del colore, mentre la luminosità descrive il livello della luce.

La valutazione del colore è la parte più stimolante della valutazione del documento perché il colore è difficile da giudicare e varia a seconda delle condizioni di visualizzazione; anche a livello scientifico, le sue proprietà sono difficili da quantificare. Comunque l'arbitro finale rimane sempre l'occhio umano.

La riproduzione del colore dipende da variabili come il livello di illuminazione al momento della cattura dell'immagine, le capacità del sistema di scansione, e la rappresentazione matematica dell'informazione del colore attraverso la catena della digitalizzazione²⁰.

Alcune caratteristiche dell'immagine digitale richiedono una loro distinzione per preservare il loro aspetto o per distinguerle da informazioni testuali sottostanti. Una pagina testuale a stampa potrebbe essere una buona candidata per una scansione bitonale; un'immagine che è macchiata o con molte annotazioni richiede una scansione in base alla scala dei grigi o a colori. La decisione se scannerizzare a colori o in bianco e nero dipenderà dall'importanza che il colore riveste nella rappresentazione del contenuto informativo degli originali. A. Kenney e O. Rieger propongono di rispondere ad alcune domande prima di compiere tale scelta²¹.

1. La riproduzione del colore è necessaria alla comprensione del documento?

In alcuni documenti essenzialmente monocromi il colore può essere stato introdotto con il tempo, l'uso, o processi impropri. Per esempio le pagine ingiallite non costituiscono una motivazione per la scansione del colore. Ma se il primo o il secondo creatore hanno introdotto il colore (es. una sottolineatura rossa su un foglio stampato), questo è considerato

²⁰ ANNE KENNEY, OYA RIEGER, *op. cit.*, p. 43.

²¹ ANNE KENNEY, OYA RIEGER, *op. cit.*

significante.

2. Qual è la natura del colore? Il colore è piatto e la palette è limitata, come nei poster, o ci sono continue forti variazioni come nelle fotografie? Maggiore è la variazione e la gamma rappresentata dal colore e più risulta difficile rendere il colore stesso.
3. Qual è la finalità del colore? In alcuni casi, come opere d'arte o fotografie, i colori dell'originale sono significativi di per sè, in altri casi lo sono per fattori sia estetici che informativi. Il colore codifica anche l'informazione: per esempio, le cartine frequentemente usano il colore per differenziare gli elementi topografici o quelli fisiografici. Talvolta, però, il significato del colore può essere rappresentato in altro modo, ha senso in questo caso conservare il colore?
4. Quanto è importante mantenere il colore? Si deve definire il grado di tolleranza per le tonalità di colore. Mentre può essere critico catturare le reali sfumature dei blu preferite da Chagall, bisogna accontentarsi di una rappresentazione digitale che approssimi il blu delle vene in un'illustrazione medica.

PROFONDITÀ E GRADAZIONE

La risoluzione del colore viene altrimenti detta "profondità" e si riferisce al numero di colori disponibili a rappresentare i colori dell'originale, ossia al numero dei possibili valori con cui ciascun pixel può essere riempito. Se prendiamo un'immagine a quattro bit è evidente che questa ha quattro "solchi", cioè quattro locazioni di memoria da riempire con una serie di 1 o 0, ovvero le quattro cifre binarie sono usate per rappresentare ciascun pixel, per cui l'immagine ha 16 valori. Se questa fosse un'immagine su una scala di grigi, ciascuna cella potrebbe essere riempita da una delle possibili gradazioni di grigio con "0000" che rappresenta il nero, e "1111" che rappresenta il bianco; di conseguenza un'immagine a 8 bit implica 256 diverse combinazioni, un'immagine a 16 bit è data da 65.536 possibili combinazioni, una a 24 bit ha oltre 16 milioni di diverse combinazioni. Il modello RGB a 24 bit è il colore standard impostato da molti scanner. Ogni pixel in

un'immagine a 24 bit è costituito da una componente rossa, una verde ed una blu. I 24 bit usano 8 bit per descrivere ogni componente rossa, blu e verde di ogni campione, offrendo 256 livelli per ognuno, quindi nel complesso descrivono l'informazione sul colore in ogni pixel e permettono una palette di quasi 17 milioni di colori.

Le variazioni della profondità dei bit permettono un controllo molto maggiore su ciò che contiene ciascuna cella o pixel sullo schermo.



Immagine 2: l'aumento della profondità offre una qualità più alta²²

Dalla profondità dipende anche la "gradazione" (shade), ossia la possibilità di accrescere il numero delle varianti dell'immagine non limitate solo al bianco o al nero, in quanto ciascun pixel può contenere molte più informazioni. Questo specifico controllo di gradazione da cella a cella produce le immagini ad alta qualità che si trovano nei progetti di digitalizzazione.

LA COMPRESSIONE

I file di immagine tendono ad essere molto grossi, ad occupare molta memoria ed a caricarsi lentamente, per questo può rendersi utile o necessario ricorrere alla loro compressione. La compressione permette di prendere i dati da un file e, usando un

²² TECHNICAL ADVISORY SERVICE FOR IMAGES, "The Digital Image" in *Technical Advisory Service for image*, Advice Paper <<http://www.tasi.ac.uk/advice/creating/pdf/image.pdf>>.

algoritmo di codifica (e decodifica), di rendere il file più piccolo. Questo algoritmo si basa su due strategie, la riduzione dell'informazione ridondante e la riduzione dell'informazione irrilevante. Nel primo caso l'algoritmo andrà a cercare ripetizioni che possono essere espresse in maniera più efficiente, quindi per esempio, dove trova 10 pixel dello stesso colore, registra l'informazione per un pixel e dichiara che i nove seguenti presentano la medesima informazione. La riduzione dell'irrelevanza punta invece a rimuovere o alterare l'informazione che non ha grande rilievo per la percezione dell'immagine, per esempio semplifica alcune informazioni sul colore. Molti formati di file forniscono una compressione "lossy", con perdita di informazione che, per quanto non necessariamente evidente all'occhio umano, implica comunque una degradazione della qualità. La compressione con perdita si basa sulla strategia della riduzione dell'irrelevanza, semplifica cioè le informazioni riguardanti l'immagine per ottenere un file fino a tre quarti o addirittura due terzi più piccolo. Nello schema che segue sono riportati i passaggi della compressione e decompressione di un'immagine con perdita di informazione.

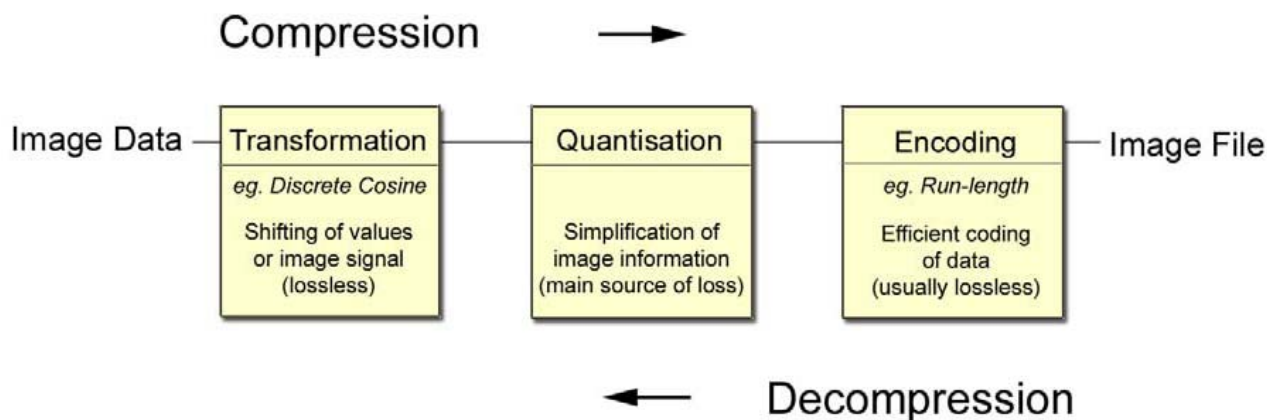
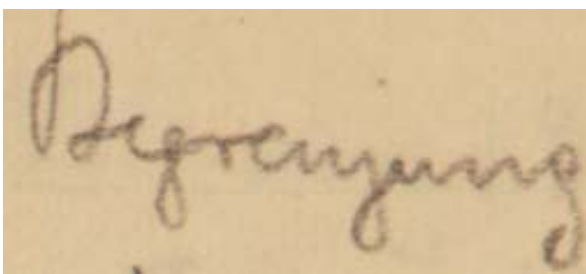


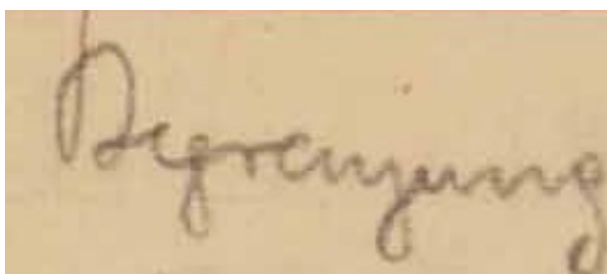
Immagine 3: fasi della compressione con perdita di un'immagine²³

Di seguito viene riportata una medesima immagine che è stata sottoposta a vari gradi di compressione.

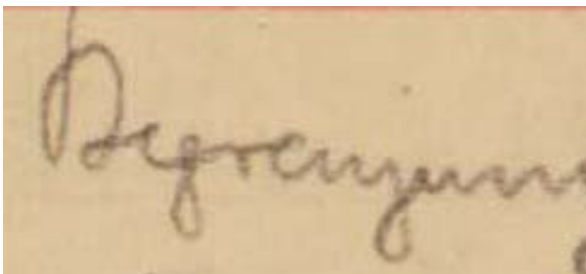
²³ TECHNICAL ADVISORY SERVICE FOR IMAGES, "File Formats and Compression" in *Technical Advisory Service for image*, Advice Paper
<<http://www.tasi.ac.uk/advice/creating/pdf/fformats.pdf>>.



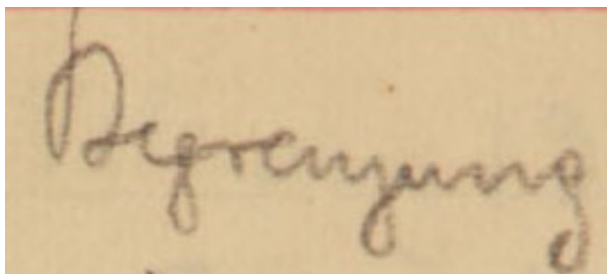
Originale in formato TIFF – 94,5 Kb



Bassa compressione – 5,85 Kb



media compressione – 5,49 Kb



alta compressione – 4,69 Kb

Immagine 4: esempi di immagini compresse

I formati che usano la compressione senza perdita sono un po' meno efficaci nella riduzione delle dimensioni delle immagini, ma permettono di mantenere la medesima qualità dell'originale non compresso. La compressione senza perdita si basa sulla tecnica della riduzione della ridondanza e si concentra sui modi più efficienti di codificare i dati riguardanti le immagini. La caratteristica più importante di questo tipo di compressione è che nessuna informazione viene mai persa, quindi l'immagine, nel momento in cui viene decompressa, apparirà esattamente come l'originale.

Esistono approcci più nuovi alla compressione, come la compressione fractals e wavelets. La loro novità è quella di non trattare più le immagini come una serie di numeri, ma rispettivamente come un'immagine e come onde continue. I formati che sfruttano questo tipo di compressione non sono però ancora molto diffusi.

I FORMATI DI IMMAGINE

File formats are orderly sequences of data used to encode digital information for storage or exchange. They are like written languages, with their own peculiar rules or grammars.²⁴

²⁴ TECHNICAL ADVISORY SERVICE FOR IMAGES, "File Formats and Compression" in *Technical*

Per quanto strutturati in maniera differente, i file di immagine generalmente cominciano con un header e sono seguiti da un body.

Ogni formato si distingue per le sue peculiari caratteristiche. Una prima caratteristica divide i formati di immagine in due grandi categorie: raster (o bit-mapped) e vettoriali (o object-oriented). Le immagini raster sono suddivise in una griglia di celle che corrispondono ad un punto (pixel) dell'immagine stessa, che ha una precisa posizione, ed a cui viene associato un colore tramite una codifica numerica. I file vettoriali invece sono un insieme di istruzioni matematiche usate da un programma di disegno per costruire l'immagine. Nella digitalizzazione si prendono solitamente in considerazione i formati appartenenti al primo gruppo in quanto il processo di cattura produce solo dati bit-mapped.

Nella pianificazione si dovrebbe tener presente quale formato di immagine occorre scegliere come standard per la cattura e per l'archiviazione dei file e quale formato di immagine occorre scegliere per l'utente finale. Le scelte sono influenzate dallo scopo del progetto e dai bisogni degli utenti che il progetto intende soddisfare.

Nella scelta si deve tenere conto dell'accessibilità e dell'utilità a lungo termine delle immagini; ciò significa che si deve scegliere uno standard commerciale o un formato non proprietario, pertanto la gamma si limita ad un ristretto numero di formati:

- Tagged Image File Format (TIFF)
- Joint Photographic Experts Group File Interchange Format (JPEG or JFIF)
- Graphic Interchange Format (GIF)
- Portable Network Graphic (PNG)

Il TIFF, Tagged Image File Format, è un formato proprietario ed è forse il più importante formato di immagine attualmente disponibile; è ampiamente usato per le piattaforme comuni e i formati di archiviazione in molti processi di digitalizzazione perché permette che immagini di alta qualità (anche superiori a 24

bit a colori) possano essere salvate senza perdita alcuna. Infine non è legato ad alcun particolare scanner o display. La conversione da TIFF ad altri formati è particolarmente agevole e molti software sono in grado di realizzarla facilmente. È diventato uno standard de facto per la creazione di immagini master ad alta qualità, offre infatti la scelta di una duplice piattaforma; generalmente però si raccomanda, per salvare i file, di usare la cosiddetta "versione PC" senza compressione.

Intorno al JPEG c'è una certa confusione poiché si tende a far coincidere la compressione JPEG con il JPEG File Interchange Format (JFIF), contraddistinto dall'estensione .jpg e da tutti indicato come "il JPEG", piuttosto che come "un'immagine compressa con una compressione JPEG"²⁵. L'algoritmo JPEG infatti, oltre al JFIF, viene usato anche negli SPIFF file (Still Picture Interchange File Format), nel TIFF e nel PDF.

I JPEG File Interchange Format sono tradizionalmente usati per visualizzare immagini a colori, supportano una profondità di 24 bit a colori e permettono anche la compressione, sebbene questo ne riduca la qualità. Questo formato è ampiamente usato per pubblicare immagini attraverso reti con una limitata ampiezza di banda, come Internet o molte intranet. Lo standard utilizza file di compressione con perdita per ridurre la dimensione dei file che devono essere trasmessi attraverso la rete. La visualizzazione dei file JPEG è supportata da tutti i browser web e da un ampio numero di applicazioni.

Il formato GIF, ossia Graphical Interchange Format, è ampiamente usato sul web, principalmente per disegni o scale di grigi, dando vita a immagini a 8 bit. Bisogna comunque notare che si tratta di un formato proprietario, coperto da licenza. Esistono due versioni del formato: 87a e 89a. Entrambi usano una caratteristica nota come "interlacing" che permette dapprima di visualizzare un'immagine a bassa risoluzione e poi gradualmente di riempirla nel dettaglio permettendo all'utente di interrompere la trasmissione in una prima fase se l'immagine non è quella desiderata. Una caratteristica simile possiedono i "JPEG progressivi", che

²⁵ TECHNICAL ADVISORY SERVICE FOR IMAGES "File Formats and Compression" in *Technical Advisory Service for image, Advice Paper*
<<http://www.tasi.ac.uk/advice/creating/pdf/fformats.pdf>>.

caricano prima un'immagine a bassa qualità, poi costruiscono la grafica che aumenta gradualmente in qualità e chiarezza.

Questi due formati costituiscono i più noti file di immagine per il web e sono di norma visualizzabili da qualsiasi browser. In base alla natura delle immagini, può essere più appropriato o il JPEG o il GIF. Il GIF è ben adattato ai cartoni, alle icone ed ai grafici più semplici, mentre il JPEG supporta meglio fotografie scannerizzate e immagini complesse.

Il formato PNG, Portable Network Graphics, è stato creato in alternativa al GIF, ha una compressione senza perdite ed offre un miglior controllo sulla luminosità dell'immagine permettendo di arrivare a 48 bit per pixel. Le immagini PNG sono supportate dalle versioni più recenti dei browser più usati.

Le immagini a bassa risoluzione sono perfettamente adeguate per la visualizzazione sul web, ma sono inadatte alla stampa o alle analisi più dettagliate, mentre le immagini ad alta qualità sono di solito troppo grandi per essere distribuite senza problemi sulla rete.

Una soluzione può essere quella dei formati di file a piramide. Questi permettono di memorizzare una gamma di risoluzioni di una particolare immagine nello stesso file: al vertice della piramide si può avere l'immagine a bassa risoluzione che può essere presentata inizialmente agli utenti come file GIF o JPEG. Gli utenti, se desiderano esplorare una sezione dell'immagine con un dettaglio maggiore, possono andare allo strato successivo della piramide. Oggi sono disponibili numerosi esempi di questo formato, fra cui FlashPix (.pfx), ma ve ne sono molti altri, i quali possono essere compressi, come ad esempio JPEG2000, che offre un approccio "a piastrelle". Questa nuova tecnologia prevede una compressione con perdita o senza perdita senza compromettere la qualità dell'immagine; la sua struttura multi-risoluzione può eliminare la necessità di avere molte versioni di diverse risoluzioni della stessa immagine in un database. Gli utenti devono però scaricare uno speciale plug-in per il loro browser.

Il Minerva working group dà delle indicazioni anche riguardo ai formati. Prima di stabilire quale formato usare bisogna prendere in considerazione gli standard più rilevanti, la piattaforma di uso consolidato a livello globale e la misura in cui i

formati dei file sono supportati dal software in uso dalla propria organizzazione e dal proprio bacino d'utenza. L'output file di default per le immagini digitali deve essere il TIFF, a meno che non ci sia una motivazione ben precisa per scegliere un altro formato; in alternativa dovrebbe comunque essere creato un formato di immagine senza perdita, cioè non compresso.²⁶

Una versione di dimensioni inferiori, in formati compressi, può essere creata usando un software di fotoritocco. La scelta dei formati deve essere governata dall'imperativo di creare l'output di più alta qualità e dalla disponibilità delle procedure di migrazione per la futura preservazione del master digitale.²⁷

Come già detto i file TIFF creati dalla digitalizzazione sono tipicamente troppo grandi e non sono quindi appropriati per la pubblicazione in Internet. Il Minerva working group consiglia quindi di creare dal file master delle versioni distribuibili che possono essere create aprendo il file TIFF in un programma di ritocco delle immagini ed esportandolo in JPEG o PNG.

La "Normativa per l'acquisizione digitale delle immagini fotografiche", emanata dall'Istituto centrale per il catalogo e la documentazione del Ministero dei beni e le attività culturali italiano, si muove in una direzione diversa, consiglia infatti l'uso del formato PNG sia per le immagini ad alta risoluzione non compresse che per le immagini a bassa risoluzione compresse²⁸.

Anche la risoluzione del colore può essere portata a 256 colori se ciò non comporta un'eccessiva perdita di dati. La scelta della giusta risoluzione del colore richiede decisioni soggettive. Bisogna stabilire quale sia la qualità "accettabile" e trovare un bilanciamento tra la qualità e la dimensione dei file. In generale il totale delle immagini in una pagina web non dovrebbe superare i 100 Kb. Immagini più grandi possono certamente essere pubblicate sul web, avvisando l'utente che l'apertura dei file potrebbe richiedere un po' di tempo.

²⁶ MINERVA WORKING GROUP, *op. cit.*

²⁷ MINERVA WORKING GROUP, *op. cit.*

²⁸ MINISTERO PER I BENI E LE ATTIVITÀ CULTURALI, ISTITUTO CENTRALE PER IL CATALOGO E LA DOCUMENTAZIONE, Normativa per l'acquisizione digitale delle immagini fotografiche, 2004
<<http://www.iccd.beniculturali.it/download/fotodig.pdf>>.

L'OTTIMIZZAZIONE DELLE IMMAGINI

Dopo che l'immagine è stata acquisita e salvata come file master, è necessario ottimizzarla per i suoi usi futuri.

Il TASI propone una distinzione del processo in due parti; inizialmente deve essere compiuta un'ottimizzazione generica di tutti i file, poi un'ottimizzazione specifica che crei le immagini surrogato per gli scopi del progetto. I passaggi proposti per ottimizzare le immagini master vengono riassunti nelle due tabelle seguenti.

Initial Generic Optimisation	The exact method of undertaking these tasks will depend upon the Image manipulation program that you choose for the job, however all stages here are quite generic and will be available in all common image manipulation programs.
Create a working copy	It is best practice to first create a working copy of the image from the master archive file, then if for any reason things go wrong within the manipulation you know that the master archive image is safe. This file should be named in such a way as to both connect it to the original and also show that it is a new working copy of the original image.
Crop if necessary	Check the size, shape and orientation of image and adjust if needed. There is no point in saving image area that is surplus to needs.
Optimise density range	Use levels tool to adjust shadow and highlight points to best use the whole range of available tones within original image.
Check and correct any fault in the colour of image file	Use curves tool to adjust and modify the colour balance within each separate channel.
Check image for any faults or artefacts	Each file should be visually checked for any marks or dust. If it is easy to fix these within software, then it should be done. If not then the image should be considered to have failed the QA and should be marked for subsequent re-capture.
Apply sharpening if necessary	It is best practise to add no sharpening at this point, however some scanners (this is less of a problem with digital cameras) have an inherent softening effect on the images and it can be necessary to apply some small element of Unsharp Masking. This should certainly be limited to only very slight work that repairs image rather than augments image.
Save your work often	Some image editors have a limited undo capability (often just the last operation). Others keep a 'History' of operations that can be selectively deleted or changed, however remember that the number of 'History' levels will quickly expand the file size and therefore memory requirements.

Tabella 2: Operazioni per l'ottimizzazione generica²⁹

²⁹ TECHNICAL ADVISORY SERVICE FOR IMAGES, "Basic Guidelines for Image Capture and

Further Specific Optimisation	After the image has been generically optimised, it will be necessary to undertake further optimisation specific to the required use of the image.
Save again	It is best practice to again save the image at this juncture so you have both the original 'Masterarchive' image and an 'Adjusted-Master' image for later use. Both file names should be unique and reflect the differing use.
Resize	Any surrogate images should be made by re-sizing direct from the 'Adjusted-Master' archive image. It is important that images are only made smaller rather than larger.
Sharpen image	If it is necessary to re-apply any sharpening, now is the time to do it, between the re-sizing and any compression made to image. Sharpening should be undertaken with great caution! It is imperative that image is not over sharpened (only to remedy rather than augment image).
Save image (within appropriate file-format)	Choose the appropriate file-type for the proposed use. This is likely to be either TIFF (for print or archive) or JPEG (for delivery on monitor). If file is being saved as a JPEG, it will be necessary to choose a level of compression for the file; this choice is a compromise between quality and filesize and will need to be deduced by testing. As a rule of thumb a JPEG compressed to 10% of its original size is normally thought to be visually acceptable in quality terms, however a larger size file can normally be compressed more than a smaller file.
Naming File	Again it is important that the file is saved using a consistent naming structure that enables you to identify both the original file and the proposed usage for the new file.

Tabella 3: Operazioni per l'ottimizzazione specifica³⁰

LE SCELTE PER IL PROGETTO

La digitalizzazione di un volume o di un documento può dare vita a un'immagine digitale della pagina oppure ad un file di testo. Nel progetto Benussi si è deciso di ricorrere al formato immagine principalmente per due motivi:

- L'immagine digitale rende conto delle caratteristiche fisiche dell'originale, come il tipo di supporto, il colore del supporto, il colore dell'inchiostro, la dimensione della pagina etc.

Optimisation" in *Technical Advisory Service for image , Advice Paper*
<<http://www.tasi.ac.uk/advice/creating/pdf/img-capt.pdf>>.

³⁰ TECHNICAL ADVISORY SERVICE FOR IMAGES, "Basic Guidelines for Image Capture and Optimisation" in *Technical Advisory Service for image , Advice Paper*
<<http://www.tasi.ac.uk/advice/creating/pdf/img-capt.pdf>>.

- Essendo i documenti nella quasi totalità manoscritti, non era possibile eseguire un OCR, cioè il riconoscimento ottico dei caratteri, ma solo una trascrizione, operazione molto più lunga e difficoltosa.

Il secondo passo è stato quello di stabilire i parametri per la cattura delle immagini ed il formato dei file. In base alla letteratura sull'argomento, e dopo aver eseguito qualche scansione di prova anche a risoluzioni superiori, è stato stabilito di compiere l'acquisizione ad una risoluzione di 300 dpi con una profondità di 24 bit. In molti progetti le grandi dimensioni del materiale hanno portato alla scelta di acquisire a diverse risoluzioni; la grandezza fisica e la forma di un documento, infatti, possono avere un forte impatto sulla risoluzione e sulla grandezza dei file. Si è comunque deciso di digitalizzare tutto il materiale alla stessa risoluzione per le caratteristiche dello stesso e per uniformità con il resto dell'archivio. Mentre per i documenti testuali 300 dpi è una buona risoluzione, in quanto risponde al requisito della leggibilità, per i dipinti e le mappe di piccole dimensioni è una risoluzione bassa, essendo però il materiale di grandi dimensioni questo livello di risoluzione sembrava un buon compromesso tra il livello di dettaglio e la dimensione dell'immagine digitale.

Altra scelta è stata quella relativa alla scansione a colori o in bianco e nero in quanto da questa dipendeva l'uso di un determinato tipo di scanner. Tale decisione, come già detto, dipende dall'importanza che il colore assume nella rappresentazione del contenuto informativo degli originali. Dal momento che Benussi prestava molta attenzione all'uso dei colori, capitava spesso che nei suoi scritti ne usasse diversi, magari per aggiungere delle note, per creare grafici o per rappresentare dati sperimentali, è sembrato naturale rispettare questa sua abitudine.

Dopo aver stabilito le caratteristiche delle immagini si è valutata la possibilità di svolgere il lavoro all'interno della biblioteca o di affidare il lavoro ad una società di servizi. La decisione finale è stata quella di affidare il lavoro di scansione in outsourcing per due motivi: l'archivio Benussi si costituisce di circa 12.000 pagine, per cui i costi per acquistare un hardware per la cattura delle immagini con le caratteristiche stabilite in precedenza non sarebbero stati ammortizzati; la

Biblioteca non aveva personale specializzato per svolgere questo tipo di attività e naturalmente l'addestramento di una persona avrebbe implicato ulteriori investimenti.

Nella scelta della società a cui affidare il lavoro la discriminante è stata la valutazione della duttilità della macchina usata nell'adattarsi alle dimensioni del materiale. Peculiarità dell'archivio sono l'eterogeneità delle dimensioni dei documenti e la presenza di alcuni fogli formato poster e di disegni molto grandi che hanno reso indispensabile l'uso di uno scanner A0 per la realizzazione del lavoro. In effetti lo scanner scelto permette di scansionare documenti di formato diverso in maniera semplice ed in modo molto veloce e permette di passare da un foglio di pochi centimetri di grandezza ad uno di 5 metri x 2 metri massimo.

La società Gallo Pomi, che ha effettuato la scansione delle immagini, usa "jumboscan", un sistema digitale prodotto dalla Lumière Technology di Parigi. Il progetto per la creazione di questo sistema fu finanziato dal museo del Louvre per indagini multispettrali, che permettessero di verificare lo stato delle tele prima del restauro. La scansione si basa sul metodo CCD monolineare³¹ ad altissima definizione, grazie anche a tre passaggi per scansione (RGB). La macchina consente una risoluzione di 20.000 x 12.000 pixel e da 150 a 1200 dpi.

Il sensore per la scansione veniva usato per alcuni satelliti di osservazione ed offre un'altissima definizione. Possiede 12.000 fotodiodi di 6.5 μ ed un'alta correzione del difetto anti-blooming. Questo livello di sensibilità, collegato ad una frequenza di campionamento di 30 bit nella modalità RGB, assicura un ampio spettro di colore. Le lenti della camera sono state sviluppate e create appositamente da un'industria tedesca con determinate caratteristiche. Il potere di risoluzione è molto alto, non solo al centro, ma anche ai lati e negli angoli; la qualità di riproduzione è costante senza perdite visibili anche con materiale molto ampio; non ha una distorsione

³¹ Il CCD, Charge Couple Device, è un sensore, ovvero un dispositivo sensibile alla luce, che la rileva in termini di segnale elettronico. Più correttamente possono essere definiti come trasduttori opto-elettronici, vale a dire dispositivi che "traducono", "trasformano" la luce in un segnale elettronico. Si presentano come piccoli rettangoli costituiti da un gran numero di celle, disposte lungo una linea sottile negli scanner e a griglia nelle macchine fotografiche. Ciascun elemento del CCD si compone di due dispositivi distinti: un fotodiodo ed un'area di trasferimento della carica (CTR). Maggiore è il numero degli elementi fotosensibili per unità di lunghezza e maggiore sarà la risoluzione finale di acquisizione.

percepibile, infatti il suo fattore E.M.Q. di distorsione non supera mai lo 0.005%. L'impianto illuminante utilizza lampade a luce fredda basculanti per illuminare esclusivamente il segmento di originale sottoposto in quel momento a scansione, per proteggere le opere da calore e da raggi U.V. Il piano d'appoggio dello scanner, di colore nero, creava qualche problema perché distorceva il colore nelle immagini; molti documenti, infatti, erano costituiti da carta velina trasparente. Per ovviare al problema sono stati usati dei fogli di cartoncino di colore neutro.

L'area di lavoro è stata approntata in uno degli uffici della biblioteca, dove, nei momenti in cui si effettuava la scannerizzazione, si è cercato di mantenere una luminosità uniforme mantenendo le luci spente e le veneziane chiuse.

Il computer usato per l'acquisizione è un Power MAC da 733 Mhz con 4 Gb di memoria RAM ed un monitor Apple Studio da 19 pollici, corredato da un unità di back up HP superstore.

La scansione dei documenti ha fornito immagini in formato TIFF, che la società ha provveduto, man mano che procedeva nel lavoro, a masterizzare su DVD. Questi DVD sono stati considerati la copia master per l'archiviazione e le immagini al loro interno non sono più state modificate, poiché si è preferito mantenere gli originali esattamente come erano stati scansionati.

Altri due computer sono stati acquistati con i fondi del progetto per procedere all'ottimizzazione delle immagini. Sono composti da una scheda madre ASUS, una CPU AMD Athlon da 2.4 Ghz, memoria RAM DDR 3 Gb, con 2 interfacce seriali ATA 133, LAN broadcom 10/100 mbps, 2 hard disk, uno per il sistema operativo ed i programmi da 20 Gb e l'altro da 60 Gb per le immagini. Sono corredati da Monitor Samsung Syncmaster da 29 pollici e da un lettore CD 52 x, due lettori DVD 16x ed 1 masterizzatore DVD. La scheda video è una NVIDIA GeForce 4MX 440 con AGP 8X, con una dimensione di 64 Mb.

Per questo tipo di lavoro è stato usato Adobe Photoshop 6.0. Ogni immagine in TIFF è stata trasferita sul computer, girata se non si presentava nel giusto verso per la lettura, ritagliata in modo da eliminare le parti non significative (l'immagine del piano d'appoggio) e compressa prima di essere salvata. Nel momento in cui veniva salvata all'immagine veniva attribuito un nuovo file name. Per la

compressione è stato usato il JPEG con un fattore di compressione medio, in modo da conciliare la necessità di avere file di dimensioni inferiori (lo spazio occupato dai file si è ridotto circa del 90%) e di mantenere comunque un grado di dettaglio tale da riuscire a leggere i documenti via web anche facendo ingrandimenti dell'immagine del 200%, operazione in alcuni casi necessaria a causa delle dimensioni dell'originale o della grafia di Benussi. Dopo il salvataggio in formato JFIF è stato necessario apportare delle piccole modifiche ad alcune immagini in cui la compressione aveva evidenziato alcuni difetti prima impercettibili. Per esempio alcune stampe su stoffa dell'amico di Benussi, Gino Parin, presenti all'interno del carteggio, a causa dell'umidità sono leggermente ammuffite. Nell'originale e nell'immagine TIFF si notano solo dei piccoli punti di colore nero o leggeri aloni giallastri, mentre nel file JPEG tutte le aree interessate sono diventate rossastre. Si è intervenuti campionando il colore dove non era stato modificato e distribuendolo nelle aree alterate. Terminato il lavoro di ottimizzazione è stata naturalmente creata una copia di back-up dei dati. In questa fase è stato anche controllato il lavoro svolto in outsourcing ed è stato necessario chiedere una nuova scansione di alcuni documenti, le cui immagini erano sfocate o presentavano delle alterazioni dovute all'illuminazione.

Per la visualizzazione sul web si era pensato, inizialmente, di creare dei file in PDF per permettere all'utente una facile visualizzazione. In seguito, viste le grandi dimensioni dei file ottenuti, si è deciso di usare il PDF come alternativa offerta all'utente per scaricare i documenti nel loro insieme. Anche dei documenti in PDF è stata creata una copia di back-up.

CAPITOLO 3: Metadati e sistemi di metadati

Affinché una biblioteca digitale possa funzionare è necessario che in essa le informazioni vengano organizzate in maniera tale che siano reperibili. Per fare ciò è necessario utilizzare i metadati.

DEFINIZIONE DI METADATO

Il significato letterale della parola metadato è “un dato su un dato”.

Antonella De Robbio ne dà la seguente definizione:

Solitamente viene inteso nel significato di “un dato strutturato sulle risorse” il quale può essere utilizzato a supporto di un’ampia gamma di operazioni tra cui la descrizione e presentazione delle risorse, la gestione delle risorse informative, e la loro conservazione a lungo termine. Il metadato è l’informazione sull’organizzazione del dato, sui vari domini del dato e sulle relazioni tra dato e suoi domini.¹

Tim Berners Lee², padre del web, lo definisce come “informazione comprensibile alla macchina su risorse web o altri oggetti”. Nella sua visione il metadato è un dato immagazzinabile all’interno della risorsa a cui si riferisce.

Martin Dillon definisce invece il metadato come “information that communicates the meaning of other information”³, mentre Roy Tennant lo definisce come “a structured description of an object or collection of objects”⁴. I metadati possono incapsulare e dichiarare le caratteristiche essenziali di una risorsa⁵.

Il termine metadato si diffonde in ambiente digitale nei primi anni novanta, quando l’aumento delle risorse in rete crea l’esigenza di un meccanismo per la

¹ ANTONELLA DE ROBBIO, “Metadati: parola chiave per l’accesso alla biblioteca ibrida”; Convegno La biblioteca ibrida, Milano, 14 – 15 marzo 2002.

² TIM BERNERS LEE, “Metadata Architecture”. January 1997.
<<http://www.w3.org/DesignIssues/Metadata.html>>.

³ MARTIN DILLON, “Metadata for Web Resources: How Metadata Works on the Web”.
<http://www.loc.gov/catdir/bibcontrol/dillon_paper.html>.

⁴ ROY TENNANT, “Metadata as if libraries depended on it” in *Library Journal*; 127 (7) 15 Apr 2002: pp.32 - 34.

⁵ TECHNICAL ADVISORY SERVICE FOR IMAGES “Metadata and Digital Images” in Technical Advisory Service for image , Advice Paper
<<http://www.tasi.ac.uk/advice/creating/pdf/metadata1.pdf>>.

loro ricerca; in realtà sembra che il termine sia stato coniato da Jack Meyers nel 1969 e che nel 1986 sia diventato un marchio registrato per opera della Metadata Company.

Con il tempo il concetto è stato esteso anche ai cataloghi delle biblioteche ed agli archivi tradizionali.

“Il sistema dei metadati è la componente logica di qualsiasi biblioteca”⁶ sia tradizionale che digitale poiché esso organizza la struttura informativa. All’interno di una biblioteca digitale l’organizzazione dei contenuti riveste un ruolo molto importante ed i metadati aiutano questa organizzazione non solo dal punto di vista catalografico, come succedeva nella biblioteca tradizionale, ma anche ad altri livelli. In ambiente digitale i metadati, oltre a descrivere la risorsa, la presentano. Le loro funzioni si moltiplicano: servono per esempio a localizzare la risorsa, a stabilirne le modalità d’accesso e a verificarne la veridicità; diventano inoltre la chiave d’accesso a diversi sistemi comunicativi, che prima non riuscivano ad esprimersi tra loro.

I VARI TIPI DI METADATI

I metadati possono essere collocati all’interno o all’esterno del documento a cui si riferiscono. Questa distinzione, fa notare la De Robbio, esisteva già prima dell’era digitale “tra la vecchia schedina bibliografica, esterna al volume, e la familiare CIP (Cataloging-In Publication) parte integrante del testo a stampa”⁷. Martin Dillon divide i metadati in due grosse aree⁸, da una parte quelli che presentano la risorsa, dall’altra i metadati descrittivi.

I metadati del primo tipo sono interni al documento, caratterizzano il contenuto della risorsa tramite tag o etichette, sono parte integrante della risorsa che descrivono e perciò sono automaticamente riconosciuti ed estratti dai sistemi di recupero ed indicizzazione.

Al secondo tipo appartengono invece gli standard per la descrizione delle risorse; questi metadati possono essere sia interni che esterni; queste

⁶ ANTONELLA DE ROBBIO, *op. cit.*

⁷ *ibidem*

⁸ MARTIN DILLON, *op. cit.*

informazioni strutturate possono infatti essere indipendenti dalla risorsa oppure essere collocate al suo interno nell'intestazione <head>.

I formati cambiano secondo la tipologia, l'implementazione che ne è stata fatta, e l'ambito di provenienza, ma l'elemento che maggiormente li differenzia è la diversa natura degli oggetti che descrivono.

Una categorizzazione può essere fornita da diversi punti di vista. Sono stati teorizzati tre tipi di approccio e tre livelli di distinzione.

1. Livello tipologico strutturale

La De Robbio, basandosi sulla teorizzazione di M. Dillon, L. Dempsey, R. Heery e M. Day, ha costruito una tabella che offre una buona visione di insieme.

Suddivisione fasce	Caratteristiche dei record	Caratteristiche associate ai formati	Formati dei record
Area dei metadati come tag= metatag			
Prima fascia	Formati semplici e non strutturati. Formati proprietari Indicizzazione su tutto il testo pieno	Localizzazione Generazione effettuata dai robot	Lycos Altavista Yahoo
Indicizzazione a testo pieno			
Area dei metadati come descrittori			
Seconda fascia	Formati strutturati Standard emergenti Strutture dei campi	Selezione Input manuale o automatico Attributi/coppie di valori	Dublin Core IAFA templates RCF1807 ROADS templates
Formati generici a struttura semplice			
Terza fascia	Formati ricchi Standards internazionali Etichettature di campo elaborate	Valutazione e analisi Creazione manuale Sottocampi	CIMI MARC TEI EAD
Strutture più complesse, domini specifici Parte di un'architettura semantica ampia			

Nella prima fascia abbiamo i metadati creati dai motori di ricerca o di formato proprietario non strutturati. L'estrazione dei tag dalla risorsa viene fatta in automatico al fine di consentire l'indicizzazione per la ricerca. Nell'indicizzazione vengono assegnate delle parole-chiave generiche, senza creare una struttura per campi. A questa fascia possono appartenere i tag delle pagine web.

Nella seconda fascia si collocano i metadati generici con una struttura semplice organizzata in campi definiti. Lo svantaggio di questi metadati è la poca dinamicità: descrivono oggetti discreti senza creare relazioni multiple. Il vantaggio è che permettono una valutazione delle potenzialità, dell'utilità e dell'interesse della risorsa in quanto orientati alla selettività dei materiali. A questa fascia appartiene, per esempio, lo standard Dublin Core.

Alla terza fascia appartengono i formati a struttura complessa. Questi metadati vengono creati manualmente da professionisti dell'informazione che compiono prima una selezione del materiale. La loro presentazione dipende dagli scopi per cui sono stati creati. Solitamente questi formati sono suddivisi in sottocampi con etichettature elaborate. A questa fascia appartengono gli standard internazionali come il MARC e l'EAD.

2. Livello funzionale

Come già accennato i metadati non hanno solo una funzione di descrizione della risorsa, ma anche di presentazione. A seconda della loro funzione possiedono caratteristiche diverse.

Tipologia	Definizione
Amministrativi	<u>Metadati utilizzati nella gestione e nell'amministrazione delle risorse informative</u> Informazioni sull'acquisizione Tracciato storico dei diritti di proprietà intellettuale, cessione, e passaggi ai fini della riproduzione Documentazione dei requisiti di accesso legale Informazioni sulla reperibilità Criteri di selezione per la digitalizzazione Controllo della versione e distinguibilità fra oggetti informativi simili Tracce di controllo create da sistemi di gestione dei metadati
Descrittivi	<u>Metadati utilizzati per descrivere o identificare risorse informative</u> Registrazioni catalografiche Indicazioni di aiuto per il reperimento Indicizzazione su database specialistici Connessione fra risorse tramite link web Annotazioni di utenti
Sulla conservazione	<u>Metadati riferiti alla gestione della conservazione delle risorse informative</u> Documentazione della condizione fisica delle risorse Documentazione delle azioni intraprese per conservare le versioni fisiche e digitali
Tecnici	<u>Metadati riferiti al funzionamento di un sistema ed al comportamento dei metadati</u> Documentazione sull'hardware e sul software Informazioni sulla digitalizzazione Tracciato storico dei tempi di risposta del sistema

	Dati di autenticazione e sicurezza
Di utilizzo	<u>Metadati riferiti al livello e al tipo di utilizzo delle risorse informativo</u> Registrazioni di visualizzazione Tracciato storico dell'uso e della gestione dei profili utenti Informazioni sulla riutilizzazione del contenuto e sull'esistenza di una pluralità di versioni

3. Livello di comunità

I metadati possono essere classificati in base all'ambito in cui vengono prodotti o impiegati.

Categoria dominio	di	Attività di dominio	Categoria dominio	di	Attività dominio
Industria		Manufatti Venditori di software Venditori di hardware Telecomunicazioni Internet Service Providers Motori di ricerca Banche e assicurazioni Pubblicità Dettaglio	Accademico		Progetti universitari e accademici Subject Gateways Iniziative W3C
Editoria		Distribuzioni Musica/multivideo distribuzione Industria registrazioni della musica Riviste scientifiche Agenzie stampa Quotidiani Gestione dei diritti	Educazione e insegnamento		Corsi interattivi Educazione permanente e continua durante l'arco della vita Insegnamento a distanza Scuola Curricula
Audio- video		Industria filmografia Produzioni multimediali	Ricerca		Laboratori di ricerca Compagnie di ricerca Società professionali
Beni culturali		Biblioteche Servizi bibliotecari Musei Portali di musei Archivi	Altri domini		Trasporti Servizi di spedizione posta elettronica Sanità Servizi governativi Ambiente

A livello internazionale esistono diversi standard⁹ di metadati elaborati da vari enti per diversi scopi.

L'applicazione dei metadati è guidata dall'uso di schemi o specifiche, consistenti in campi predefiniti per specifici tipi di informazioni. I Metadata elements sono i

⁹ Uno standard è un insieme di regole per la comunicazione tra gli uomini, mentre un formato è un insieme di regole per la comunicazione tra le macchine.

singoli componenti che vanno a formare uno schema. Ogni elemento conterrà un particolare tipo o una particolare categoria di informazioni, che dipenderanno dalla definizione degli elementi stessi. Per esempio, molti schemi contengono l'elemento Title, che riassume il contenuto dell'oggetto descrittola descrivere, e l'elemento Creator che comprende il nome della persona o dell'ente che ha creato o ha dato origine all'oggetto. Non tutti gli schemi includono gli stessi elementi: questi variano in base alle specifiche e in base alle necessità delle differenti comunità, di conseguenza variano anche i tipi e le categorie di informazioni che saranno importanti per gli utenti¹⁰.

II Dublin Core

Il Dublin Core si è ormai imposto come il principale standard di metadati nel mondo di Internet. Il suo scopo è l'accesso alle risorse digitali ed esso costituisce un compromesso tra l'indicizzazione minima che viene data alle risorse Internet automaticamente dai motori di ricerca e la vera e propria catalogazione, effettuata con i formati bibliografici, troppo costosa però per un volume così esteso di documenti.

Inizialmente il DC era stato pensato per permettere ai creatori di pagine web di autocatalogare le risorse che immettevano in rete; negli ultimi anni le sue finalità si sono ampliate ed è stato implementato uno strumento

che promuova un nuovo approccio integrato alle informazioni, che venga incontro alle esigenze dell'utenza, che offra servizi e tecnologie a tutte le attività e che tra le sue strategie stabilisca un consenso necessario per la standardizzazione a tutti i livelli.¹¹

Nel 1995 OCLC organizzò un joint workshop a Dublin, in Ohio, per discutere le semantiche dei metadati. La discussione portò all'elaborazione del primo set di 15 elementi, il "Dublin Core Metadata", ed alla costituzione della Dublin Core Metadata Initiative, un'organizzazione che si occupa dell'adozione di metadati interoperabili e dello sviluppo di vocabolari specializzati per la descrizione delle

¹⁰ TECHNICAL ADVISORY SERVICE FOR IMAGES "Metadata and Digital Images" in *Technical Advisory Service for image , Advice Paper*
<<http://www.tasi.ac.uk/advice/creating/pdf/metadata1.pdf>>.

¹¹ Dublin Core Metadata Element Set, versione 1.1: traduzione italiana. Disponibile all'indirizzo
<<http://www.iccu.sbn.it/dublinco.html>>.

risorse che permettano la creazione di più intelligenti sistemi di recupero dell'informazione.

La lingua di sviluppo del DC standard è l'inglese, ma dal 2001 si è riconosciuta la necessità che la lingua dei metadati sia la stessa lingua della risorsa descritta; per questo DCMI ha sostenuto lo sviluppo di un'infrastruttura che fornisca definizioni e documentazione riguardante i metadati nelle lingue dei suoi utilizzatori, che oggi sono più di 20.

Il set di 15 elementi è rimasto invariato dopo lo sviluppo, ma nella versione 1.1 è stato ampliato con l'aggiunta di 10 attributi. Quest'ultima versione è stata votata dall'International Standard Organization (ISO) come DIS 15386.

È riportato qui di seguito uno schema basato sulla traduzione italiana dell'element set, effettuata dall'ICCU:

NOME ELEMENTO	IDENTIFICATORE	DEFINIZIONE
Titolo	Title	Un nome, formalmente conosciuto, dato alla risorsa.
Creatore	Creator	Un'entità che ha la responsabilità principale della produzione del contenuto della risorsa.
Soggetto	Subject	L'argomento della risorsa.
Descrizione	Description	Una spiegazione del contenuto della risorsa. Per esempio un riassunto analitico o un indice.
Editore	Publisher	Un'entità responsabile della produzione della risorsa, disponibile nella sua forma presente.
Autore di contributo subordinato	Contributor	Un'entità responsabile della produzione di un contributo al contenuto della risorsa.
Data	Date	Una data associata ad un evento del ciclo di vita della risorsa, normalmente quella di creazione o alla disponibilità.
Tipo	Type	La natura o il genere del contenuto della risorsa. Il valore deve essere attribuito da un vocabolario controllato (ad esempio, la lista di lavoro dei Dublin Core Types [DCT1]).
Formato	Format	La manifestazione fisica o digitale della risorsa. Può includere il tipo di supporto o

		le dimensioni della risorsa. Format può essere usato per determinare il software, l'hardware o altro apparato necessario alla visualizzazione o all'elaborazione della risorsa.
Identificatore	Identifier	Un riferimento univoco alla risorsa nell'ambito di un dato contesto. Viene consigliato l'uso di uno di questi sistemi: l'Uniform Resource Identifier (URL) (incluso l'Uniform Resource Locator (URN)), il Digital Object Identifier (DOI) e l'International Standard Book Number.
Fonte	Source	Un riferimento ad una risorsa dalla quale è derivata la risorsa in oggetto.
Lingua	Language	La lingua del contenuto intellettuale della risorsa.
Relazione	Relation	Un riferimento alla risorsa correlata.
Copertura	Coverage	L'estensione o scopo del contenuto della risorsa. Include la localizzazione spaziale (il nome di un luogo o le coordinate geografiche), il periodo temporale (l'indicazione di un periodo, una data o un range di date) o una giurisdizione (ad esempio il nome di un'entità amministrativa).
Gestione dei diritti	Rights	Informazione sui diritti esercitati sulla risorsa.

Le principali caratteristiche di questa versione sono:

- La semplicità d'uso: anche gli autori sono in grado di capire ed usare il set di elementi, non solo gli specialisti.
- L'interoperabilità semantica grazie alla quale è possibile costituire una rete comune di dati per significato e valore.
- La pubblicità: il set si propone come strumento per la creazione di un'infrastruttura a livello internazionale.
- La flessibilità che rende possibile l'integrazione e lo sviluppo di una struttura di dati con diversi significati semantici in base al contesto della risorsa.

Ad ogni elemento, come già accennato, sono stati associati, in base alla norma ISO 11179, i 10 attributi seguenti.

Nome – Etichetta assegnata al dato

Identificatore – Identificativo univoco assegnato al dato

Versione – Versione del dato

Registrazione di autorità – Entità autorizzata a registrare il dato

Lingua – Lingua nella quale il dato è indicato

Definizione – Indicazione che rappresenta chiaramente il concetto e la natura essenziale del dato.

Obbligatorietà – Indica se il dato è richiesto sempre o solo in alcuni casi (contiene un valore)

Tipo di dato – Indica la tipologia del dato che può essere rappresentata nel valore del dato stesso

Occorrenza massima – Indica un limite alla ripetitività del dato

Commento – È un'osservazione che concerne l'applicazione del dato.

Il Dublin Core si adatta bene alla descrizione generica delle risorse o a risorse eterogenee tra loro, meno alla descrizione di informazioni strutturate gerarchicamente come i metadati riguardanti un archivio, in quanto non consente la creazione di rimandi tra i vari documenti. Inoltre, benché sia stato concepito come referente comune per diverse comunità, usa una terminologia che non si adatta perfettamente ad alcuni contesti e nella maggior parte dei casi deve essere completato da altri schemi di metadati.

Il DMCI sta cercando di sviluppare delle specifiche di questo metadato, compatibili con RDF e XML in cooperazione con il W3C¹², mentre i suoi elementi possono già essere codificati in metatag HTML.

¹² Il World Wide Web Consortium <<http://www.w3c.org>> è un consorzio che sviluppa tecnologie (specifiche, linee guida, software, e strumenti) per portare il Web al massimo del suo potenziale, definendo protocolli comuni che ne favoriscano l'evoluzione e assicurino l'interoperabilità. Il W3C, guidato da Tim Berners-Lee, Direttore e creatore del World Wide Web, è stato costituito nell'ottobre 1994 con lo scopo di sviluppare al massimo il potenziale del World Wide Web. È un consorzio internazionale di imprese, neutrale rispetto ai venditori, ospitato congiuntamente da Massachusetts Institute of Technology Laboratory for Computer Science (MIT/LCS) negli Stati Uniti, Institut National de Recherche en Informatique et en Automatique (INRIA) in Europa, Keio University Shonan Fujisawa Campus in Giappone.

L'eXtensible Mark-up Language

L'XML¹³ è un linguaggio di mark-up, cioè un sistema di istruzioni che permette di descrivere strutture testuali a vari livelli usando comandi specifici. È formato da testo e da istruzioni, i tag o etichette, non molto diverse dai simboli usati dai correttori di bozze, chiamati appunto mark-up. Per il programma che la legge ogni istruzione corrisponde ad un'azione. Sulla base del tipo di significato dei tag esistono 2 tipi di linguaggio, quelli procedurali e quelli dichiarativi. L'XML è un linguaggio di tipo dichiarativo: la sua semantica predica l'appartenenza di una porzione di testo ad una certa classe di caratteri testuali, permette cioè di associare a porzioni di testo determinate caratteristiche descrivendo la struttura del testo. Le caratteristiche strutturali vengono descritte mediante un insieme di marcatori inseriti nel testo. Ogni istruzione è una dichiarazione della forma, che asserisce l'appartenenza di una certa porzione di testo ad una determinata classe di caratteri testuali. Il linguaggio dichiarativo utilizza codifiche analitiche o descrittive, codifica cioè strutture astratte come quelle editoriali, quelle morfo-sintattiche e quelle semantiche. I linguaggi di mark-up procedurali, come l'HTML, si prestano invece a codificare le caratteristiche rappresentazionali, come la struttura compositiva, la resa grafica e la formattazione, cioè a produrre un determinato output su schermo.

L'XML non ha avuto una sua realizzazione autonoma, è stato creato come una forma ristretta del SGML che risultasse più facilmente portabile tramite Internet. È stato sviluppato nel 1996 dall'XML Working Group con il patrocinio del W3C. Le caratteristiche, che hanno permesso la grande diffusione di questo linguaggio, sono:

- Standardizzazione.
- Metalinguaggio. La qualità di essere un linguaggio che descrive altri linguaggi di mark-up.
- Generalizzazione. È un linguaggio dichiarativo altamente flessibile, in grado di descrivere vari livelli strutturali di un documento.
- Indipendenza dal sistema informatico. È autonomo dalla piattaforma hardware e software.

¹³ Informazioni relative allo standard si trovano sul sito del W3C all'indirizzo <<http://www.w3.org/XML>>.

- Pubblicità. Tutti i software conformi a XML sono in grado di gestire dati in questo formato.
- Indipendenza dai dispositivi. Un documento in questo linguaggio può essere archiviato su qualsiasi supporto digitale e visualizzato con qualsiasi dispositivo di output (on-line, su cd-rom, su carta etc).
- Indipendenza dalle applicazioni. Può rappresentare qualsiasi tipo di documento e carattere testuale, indipendentemente dalle finalità applicative.
- Indipendenza dal set di caratteri. I testi possono essere redatti con qualsiasi sistema alfabetico.

Ogni documento XML è formato da entità, unità di memoria che contengono dati analizzati o non analizzati. I dati sono costituiti da caratteri testuali che si dividono in character data, cioè il contenuto, e in mark-up, cioè le etichette che identificano l'entità. I marcatori descrivono l'organizzazione della memorizzazione e la struttura logica del documento, quindi la sua struttura fisica e logica. Ogni documento inizia con una dichiarazione della versione che si sta utilizzando e della Document Type Definition (DTD). La DTD è un elenco di dichiarazioni espresse nella sintassi XML che descrivono la struttura del documento. Le dichiarazioni definiscono gli elementi strutturali del documento, mediante un identificatore generico, il modello del contenuto di ogni elemento (content model) ed i rapporti tra i due. Ad ogni elemento possono essere associati una serie di attributi che ne specificano ulteriori caratteristiche o funzioni non strutturali. Un documento XML risulta valido solo se viene associato ad una DTD e se ne rispetta i vincoli.

L'Encoded Archival Description

L'EAD¹⁴ è una specifica DTD del linguaggio SGML, nata per iniziativa della biblioteca dell'Università di Berkeley nel 1993. Lo scopo del progetto avviato a Berkeley era quello di sviluppare uno standard non proprietario che riproducesse strumenti di corredo, come inventari, cataloghi, indici e altri

¹⁴ Informazioni sullo Standard sono reperibili sul sito della Library of Congress all'indirizzo <<http://www.loc.gov/ead/eaddsgn.html>>, inoltre sull'argomento è interessante il seguente articolo DANIEL PITTI, "Encoded Archival Description: An Introduction and Overview", in *D-Lib Magazine*, 5 (11) 1999.

documenti creati da archivi, biblioteche e musei, in un formato leggibile dal computer per l'uso del loro posseduto. Era stato sviluppato uno standard che avesse i seguenti requisiti:

- capacità di presentare descrizioni ampie e correlate come negli inventari;
- capacità di conservare le relazioni gerarchiche tra i vari livelli di descrizione;
- capacità di presentare informazioni che passino da un livello gerarchico ad un altro;
- capacità di muoversi all'interno di una struttura gerarchica di informazioni e di supportare l'indicizzazione e la ricerca di elementi specifici.

Con questi principi il gruppo di lavoro ha proceduto a rivedere la struttura del documento da codificare. A livello più semplice, un inventario consiste di due segmenti di cui uno fornisce informazioni riguardo all'inventario in sé (il suo titolo, il compilatore, la data di compilazione), l'altro informazioni riguardo al materiale d'archivio. Seguendo l'esempio della Text Encoding Initiative (TEI), il gruppo di lavoro ha designato il primo segmento come "header".

Le informazioni sul materiale descritto possono essere rappresentate in due modi:

- 1) gerarchicamente organizzate, così che l'informazione descriva un'unità di records o carte insieme con le sue parti componenti o divisioni.
- 2) Informazione aggiuntiva che non descriva direttamente i records o le carte ma che faciliti il loro uso ai ricercatori.

La gerarchia delle informazioni, che riflette i principi dell'organizzazione archivistica, generalmente inizia con un riassunto dell'insieme ricavato dalla delineazione delle parti come un set di visualizzazioni contestuali. Le descrizioni delle parti ereditano le informazioni dalle descrizioni d'insieme.

L'accordo su questa struttura complessiva ha permesso agli sviluppatori di valutare gli elementi che hanno codificato nel modello BFAP. Questi elementi che sono sopravvissuti al processo di valutazione formavano due categorie: gli elementi che potrebbero essere etichettati nello specifico, punti prevedibili nella descrizione delle unità o delle parti componenti (elementi descrittivi), e gli elementi che potrebbero essere etichettati ovunque all'interno del documento

(elementi generici). Gli elementi generici sono di solito inseriti all'interno dell'elemento descrittivo.

Una caratteristica del SGML è la possibilità di definire gli attributi e di associarli con particolari elementi. Gli sviluppatori della EAD hanno concluso che una DTD avrebbe portato grandi vantaggi, in quanto gli attributi avrebbero potuto fornire le opzioni per rendere un elemento più specifico e in tal modo un piccolo set di elementi di base avrebbe potuto essere dilatato attraverso gli attributi, invece di creare un ampio set di elementi specifici.

Combinando gli elementi descrittivi e quelli generici con gli attributi nella struttura semplificata di un documento, il team della Bentley è stato in grado di estrapolare il basilare finding aid tag set ed ha deciso:

- la separazione dell'informazione riguardo ai finding aid;
- la distinzione tra l'informazione aggiunta e l'informazione dell'unità di descrizione presentata gerarchicamente.

Il gruppo che lavora sulla EAD DTD enfatizza l'importanza della documentazione, come i tag library e le linee guida per l'applicazione, per rendere l'implementazione del SGML fattibile. Una documentazione di questo tipo necessita di essere abbastanza accessibile da permettere agli utenti di conoscere l'SGML e di applicarne la DTD sia abitualmente che saltuariamente nel loro lavoro. Il team ha fatto supposizioni riguardo alle future possibilità: allegare uno script per offrire aiuto on line e spiegare le pratiche di descrizione, collegare le informazioni ai glossari centrali e condividere le storie amministrative, presentare nuove visualizzazioni che potrebbero trasformare dati gerarchici in alberi genealogici.

II Resource Descriptive Format

RDF è lo strumento base per la codifica, lo scambio e il riutilizzo di metadati strutturati e consente l'interoperabilità tra applicazioni che si scambiano sul Web informazioni machine-understandable¹⁵. Si tratta di uno standard ancora più generico del DC, in quanto consente ai dati in questo formato di entrare direttamente nel DC o nel XML.

¹⁵ ORESTE SIGNORE, "RDF per la rappresentazione della conoscenza", <<http://www.w3c.it/papers/RDF.pdf>>.

Lo sviluppo di RDF come un framework generale per i metadati – e come meccanismo generale di rappresentazione della conoscenza per il web – è stato fortemente ispirato da PICS, Platform for Internet Content Selection, un meccanismo per il communicating rating delle pagine web dal server ai clienti. Questi indici, o *rating labels*, contengono informazioni riguardo ai contenuti delle pagine web. Lo sviluppo di PICS era motivato dalla necessità di anticipare le restrizioni su Internet proposte dalla recente legislazione americana. RDF è una fondazione per l'elaborazione dei metadati e consente l'interoperabilità tra applicazioni che scambiano informazioni, comprensibili dalla macchina, sul web. RDF dà rilievo alle strutture per permettere il trattamento automatizzato delle risorse web. I suoi metadati possono essere usati in molteplici aree di applicazione: nella *ricerca delle risorse* fornisce migliori prestazioni dei motori di ricerca; nella *catalogazione* per la descrizione dei contenuti e delle relazioni fra i contenuti stessi disponibili in un particolare sito o pagina o biblioteca digitale; nei *software agents* per facilitare la condivisione e lo scambio di conoscenza; nell'indice dei contenuti, per descrivere collezioni di pagine che rappresentano un singolo documento logico; per indicare i diritti di proprietà intellettuale delle pagine web. RDF associato alle firme digitali sarà la chiave per costruire il "Web of Trust".

Lo sviluppo di RDF è stato motivato in particolare dai seguenti usi:

- Metadati per il web. Fornisce informazioni riguardo alle risorse web ed ai sistemi che le usano.
- Applicazioni che richiedono modelli di informazione aperti (come l'organizzazione delle attività, la descrizione di processi organizzativi, l'annotazione di risorse web, etc.).
- Possibilità di elaborare i dati al di fuori del particolare ambiente in cui sono stati creati.
- Interworking tra applicazioni: combinando dati di diverse applicazioni si arriva ad una nuova informazione.
- Elaborazione automatica delle informazioni ipertestuali dai software agents.

In senso esteso il fine del RDF è definire un meccanismo per la descrizione delle risorse che non si basi su particolari domini dell'applicazione, o che definisca la semantica di ogni dominio. Si tenta di creare un meccanismo neutrale, affinché sia adattabile a qualsiasi dominio. Questo scopo generale sarà raggiunto fissando dei traguardi intermedi¹⁶:

- avere un semplice data model
- avere una semantica formale
- usare un vocabolario estensibile basato su un URI
- usare una sintassi basata sul XML
- supportare l'uso di informazioni basate sullo schema XML

RDF non compie una descrizione semantica delle risorse, ma costituisce la base comune per poterla esprimere con metadati diversi che possano scambiare informazioni tra loro.

Il framework è composto da due elementi, l'RDF Model and Syntax, che stabilisce il data model e codifica l'XML, e l'RDF schema che implementa vocabolari specifici per i metadati.

La sintassi è l'organizzazione semantica del data element per l'elaborazione autonoma dei metadati e fornisce un modello di dati basato su un semplice diagramma in cui la rappresentazione delle informazioni è poco vincolante e flessibile, così da facilitare lo scambio e l'uso dei metadati tra diverse applicazioni.

Questa sintassi può descrivere e definire solo risorse identificabili attraverso un URI (Uniform Resource Identifier), composto da tre tipi di oggetto:

- RISORSA, vale a dire qualunque cosa possa essere descritta da un'espressione RDF.
- PROPRIETÀ, cioè un aspetto specifico, una caratteristica, un attributo associati ad un nome che assumono nell'insieme determinati valori.
- STATEMENT, costituito dalla risorsa contraddistinta da un nome e da un valore propri che nell'insieme formano una tripla.

¹⁶ GRAHAM KLYNE, JEREMY CARROLL, "Resource Description Framework (RDF): Concepts and Abstract Syntax", <http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/>>.

La tripla è la struttura di fondo di ogni data model, è formata da un soggetto, un predicato ed un nome e viene solitamente rappresentata come un grafo etichettato orientato.



Fig.1: Rappresentazione grafica di un data model RDF

Il grafo rappresenta la relazione della tripla come un collegamento nodo – arco – nodo. I nodi rappresentano il soggetto e l'oggetto, mentre l'arco rappresenta il predicato o la proprietà e per questo punta sempre verso l'oggetto. Questo tipo di struttura permette la creazione di risorse a più livelli adattandosi quindi a raccolte gerarchiche come gli archivi.

Per poter fare riferimento a più di una risorsa sono stati creati tre tipi di contenitori:

- Bag, una lista non ordinata di risorse o costanti, che dichiara i valori multipli di una proprietà senza che l'ordine dei valori sia significativo.
- Sequenze, una lista ordinata in cui l'ordine dei valori multipli acquista un preciso significato.
- Alternative, una lista di risorse o costanti alternative per valori di singole proprietà.

Grazie ai questi contenitori una risorsa potrà avere un soggetto con più dichiarazioni rette dal medesimo predicato.

RDF in sé non contiene nessun vocabolario predefinito di authoring metadata; usa riferimenti URI per identificare risorse e proprietà. RDF fornisce significati specifici ad alcuni riferimenti URI. In particolare, riferimenti URI con la seguente sottoserie invariabile sono definiti dalle specifiche RDF:

- <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
(convenzionalmente associata con il prefisso namespace `rdf:`)

I termini del vocabolario in rdf sono elencati nella RDF syntax specification ([RDF-SYNTAX](#)). Alcuni di questi termini sono definiti dalle specifiche RDF per

indicare concetti determinati, altri hanno scopi sintattici (es. rdf:ID è parte della sintassi RDF/XML).

Ci si deve, comunque, aspettare che saranno compilati dei vocabolari per gli standard poiché è un'esigenza essenziale per l'interoperabilità su larga scala. Chiunque può creare un nuovo vocabolario; l'unico requisito per usarlo è che un URI sia incluso negli esempi di metadati.

LA SCELTA ALL'INTERNO DEL PROGETTO BENUSSI

Nello scegliere il livello di complessità dei metadati si è fatta una valutazione in base ai costi ed alla facilità di creazione e mantenimento. Il TASI, Technical Advisory Service for image, nelle sue best practices afferma:

descrivere ogni singola immagine è molto più costoso e richiede un lavoro maggiore che descrivere una raccolta di immagini, [...] assegnare termini da vocabolari controllati e schemi di classificazione è più costoso e impegnativo che assegnare parole chiave in modo casuale. Il livello di complessità dei metadati usati dovrebbe essere in gran parte dettato dai bisogni degli utenti della collezione. Il costo della creazione e della manutenzione di una risorsa sofisticata multifacettata con ricchi strati di descrizione può essere bilanciato fino ad un certo punto se questo scenario può essere mostrato alle esigenze degli utenti conosciuti. Un'approfondita comprensione della potenziale comunità di utenti della collezione attraverso un'investigazione preliminare dei bisogni dell'utente è una componente essenziale di ogni progetto di digitalizzazione.¹⁷

Sostiene inoltre che:

La decisione, su quale schema di metadati adottare e quali livelli di descrizione applicare, dovrebbe essere guidata sia dai bisogni che dalla comunità degli utenti oltre che dai sistemi esistenti e dai protocolli usati e promossi dalle organizzazioni che offrono servizi simili o accesso alle risorse.

Si è deciso di esplicitare le informazioni relative ai file di immagine una sola volta nella pagina di presentazione dell'archivio, dal momento che queste si ripetevano. La scelta di usare l'UNIMARC per la descrizione dell'archivio, che

¹⁷TECHNICAL ADVISORY SERVICE FOR IMAGES "Metadata and Digital Images" in Technical Advisory Service for image , Advice Paper
<<http://www.tasi.ac.uk/advice/creating/pdf/metadata1.pdf>>.

può sembrare un po' azzardata, è derivata dalla decisione di usare il sistema di gestione già in possesso della Biblioteca dell'Università.

Tale scelta è stata operata, oltre che per una questione di economia, per i seguenti motivi:

- la creazione di un'integrazione tra le risorse presenti nella biblioteca stessa;
- la possibilità di trasformare automaticamente, in un secondo momento, le descrizioni in formato XML;
- la rispondenza del formato UNIMARC ai requisiti funzionali necessari nel database.

Marieke Guy, Andy Powell e Michael Day, nel loro articolo intitolato "Improving the Quality of Metadata in Eprint Archives"¹⁸, propongono una lista di domande per capire se i metadati possono essere definiti di buona qualità e se i requisiti funzionali sono osservati:

Functional Requirements List

We would like users to be able to:
Search records by:

Title
Keyword
Year

Browse records by:

Title
Year
Location?

View latest additions to the archive

We would like to be able:

Link together description and digital object.

¹⁸ MARIEKE GUY, ANDY POWELL, MICHAEL DAY, "Improving the Quality of Metadata in Eprint Archives", in Ariadne, (38). <<http://www.ariadne.ac.uk/issue38/guy>>.

La conclusione a cui si giunge nell'articolo è che la qualità dei metadati può essere ritenuta soddisfacente, se le loro caratteristiche supportano i requisiti funzionali definiti in precedenza. Il formato UNIMARC, all'interno del progetto, rispondeva alle esigenze definite dalla Functional Requirements List sopra riportata, quindi è stato adottato, anche se con qualche adattamento agli standard di descrizione archivistica (ISAD).

L'UNIMARC

UNIMARC¹⁹ è un formato di dati, cioè una struttura logica di dati comprensibile per la macchina. I formati possono essere di due tipi, interni alla macchina (cioè interni al software per esigenze di gestione del sistema locale) oppure usati per lo scambio dei dati fra sistemi diversi. A questa seconda categoria appartiene l'UNIMARC.

Un formato è costituito da tre elementi:

1. Struttura fisica che stabilisce le regole di costruzione delle registrazioni, come la tipologia dei campi (lunghezza fissa o variabile) e la loro ripetibilità.
2. Identificatori del contenuto che identificano e distinguono gli elementi. Le etichette sono costituite da una stringa di tre caratteri che identifica il tipo di campo; gli indicatori, frapposti tra etichette e dati, forniscono informazioni sui contenuti dei campi e infatti assumono valori diversi in base ai campi in cui vengono utilizzati. Infine i codici di sottocampo si trovano solitamente all'interno dei campi ed introducono gli elementi dei sottocampi (si tratta solitamente di lettere minuscole associate al delimitatore \$).
3. Contenuto della registrazione, cioè l'insieme dei dati contenuti nei campi e nei sottocampi.

A metà degli anni '60 la Library of Congress sentì l'esigenza di trasferire la catalogazione dal formato cartaceo ad un formato leggibile dall'elaboratore, fu creato così un apposito working group che elaborò nel 1964 il formato MARC I,

¹⁹ Informazioni su questo formato sono reperibili sul sito dell'IFLA all'indirizzo <<http://www.ifla.org/VI/3/p1996-1/sec-uni.htm>>, in lingua italiana uno dei maggiori contributi sull'argomento viene dal manuale ANTONIO SCOLARI, *UNIMARC*, Milano, Bibliografica, 2000.

sperimentato per alcuni anni da un consorzio di biblioteche statunitensi. Nel 1968, in collaborazione con la British Library, furono apportate delle modifiche al formato che assunse il nome di MARC II. Successivamente negli anni '70 furono elaborati, a livello nazionale, diversi formati che facevano riferimento al MARC, tra cui INTERMARC in Francia, USMARC negli USA, UK MARC in Gran Bretagna e ANNAMARC in Italia. Nel 1971 si pensò all'elaborazione di un formato internazionale, SUPERMARC, che permettesse la comunicazione dei vari MARC nazionali. In corso di elaborazione si pensò all'adozione dello standard ISBD per la descrizione bibliografica e la creazione di un formato unico per tutti i tipi di materiale. Nel '75 uscì la prima bozza di UNIMARC, mentre nel 1977 fu stampata la prima edizione ufficiale.

Il formato è organizzato in blocchi funzionali, contraddistinti da etichette con una logica precisa:

0	Blocco di identificazione
1	Blocco delle informazioni codificate
2	Blocco delle informazioni descrittive
3	Blocco delle note
4	Blocco dei legami
5	Blocco dei titoli in relazione
6	Blocco dell'analisi semantica
7	Blocco della responsabilità intellettuale
8	Blocco dei dati internazionali
9	Blocco di uso locale

Caratteristiche di questo formato sono la grande analiticità, l'ospitalità verso le diverse esigenze di catalogazione e l'enfasi sui legami bibliografici.

All'inizio di ogni record bibliografico è obbligatoriamente presente il campo "guida" (record label) che aiuta l'elaboratore a decodificare correttamente la registrazione. Non ha una sua etichetta ed è formato da 24 caratteri (da 0 a 23), di cui alcuni hanno valori fissi ed altri che variano in base alla versione. I valori per la maggior parte delle posizioni sono calcolati dal computer, mentre le posizioni 6 (la tipologia del materiale), 7 (il livello bibliografico), 8 (il livello

gerarchico), 17 (il livello di codifica) e 18 (la forma della descrizione) devono essere compilate dal catalogatore.

Una guida di UNIMARC è quindi così composta:

Posizioni	0 – 4	5	6 – 9	10	11	12 – 16	17 – 19
	20 – 23						
Codici	00667	n	am0#	2	2	00241	###
	450#						

Esempio di guida di un record inserito in ALEPH

Le posizioni da 0 a 4 indicano la lunghezza della registrazione, la posizione 5 indica lo stato di registrazione, la posizione 6 indica il tipo di registrazione, la 7 indica il livello bibliografico, la 8 il livello gerarchico, la posizione 9, come la 19 e 23 non sono definite e sono contraddistinte dal simbolo #. La posizione 10 riguarda la lunghezza degli indicatori, mentre la posizione 11 indica la lunghezza dei codici di sottocampo; queste due posizioni sono sempre contraddistinte dal numero 2. Le posizioni dalla 12 alla 16 indicano la posizione di partenza dei dati, la posizione 17 contraddistingue il livello di qualifica di UNIMARC, mentre la 18 il tipo di catalogazione descrittiva. Anche le ultime quattro posizioni presentano valori costanti. La posizione 20 indica la lunghezza della "lunghezza di campo", la 21 la lunghezza della "posizione di inizio campo" e la 22 la lunghezza della parte di implementazione.

Per la descrizione del Fondo Benussi sono stati utilizzati solo sette campi facenti parte dei primi 4 blocchi funzionali, quelli in cui vengono riportati i dati di descrizione bibliografica.

Il campo 001, il numero di registrazione, è un campo obbligatorio attribuito automaticamente dal sistema nel momento in cui viene creato un nuovo record, in questo modo ogni record bibliografico ha un suo numero di sistema che lo identifica univocamente all'interno del database e che consente di richiamarlo direttamente. Il campo 101 descrive la lingua di pubblicazione; nel campo 200 \$a si inserisce il titolo proprio di un'opera. Se questo sottocampo è seguito dall'indicatore "1", come nel caso presente, diventa il punto di accesso privilegiato del record. Il campo 210, seguito dal sottocampo \$d, serve ad inserire la data di pubblicazione; il campo 215 è quello in cui viene inserita la

descrizione fisica del documento, intesa come estensione e dimensione del materiale, per questo il campo è stato associato all'elemento "Consistenza dell'unità di descrizione" delle ISAD. Il campo 300 è quello in cui devono essere inserite le note generali, mentre il campo 307 è quello in cui vengono inserite le note relative alla descrizione fisica del materiale, questo quindi è stato usato per descrivere l'altro elemento delle ISAD riferito all'aspetto fisico dei documenti, le "Caratteristiche materiali".

Le ISAD(G)

L'international Standard Archival Description è un tipo di descrizione multilivellare. La descrizione archivistica può avere diversi gradi di analiticità: si può decidere di descrivere solo il fondo a livello generale, le singole serie, i singoli fascicoli o addirittura i singoli documenti. Si deve sempre partire dal livello più generale e scendere al particolare; ad ogni livello è obbligatorio fornire alcune informazioni. Solo in casi particolari si arriva a descrivere i singoli documenti e solitamente il fascicolo è ritenuto l'unità di descrizione archivistica.

La prima bozza è stata elaborata agli inizi degli anni novanta dall'ICA Ad Hoc Commission on Descriptive Standards (ICA/DDS), mentre la redazione ufficiale è stata approvata a Madrid nel 1992. In seguito ogni nazione ha provveduto alla traduzione nella propria lingua. In Italia la traduzione, approntata da un gruppo di lavoro operante all'interno dell'Ufficio centrale per i beni archivistici, ha richiesto quasi 2 anni di lavoro ed è uscita nel 1994. Nel 2000 la Commissione, ormai diventata organo permanente dell'ICA, ha finito la revisione dello standard approntando la seconda edizione.

Nell'introduzione²⁰ delle ISAD(G) leggiamo le opportunità offerte dall'uso di questo standard:

- a. assicurare la creazione di descrizioni coerenti, appropriate e che si spiegano da sè
- b. facilitare il reperimento e lo scambio di informazioni riguardo materiale archivistico;

²⁰ ISAD(G) : general international standard archival description : adopted by the Committee on Descriptive Standards, Stockholm, Sweden, 19-22 September 1999.
<http://www.ica.org/biblio/cds/isad_g_2e.pdf>.

- c. consentire la condivisione dell' authority data;
- d. rendere possibile l'integrazione di descrizioni da differenti localizzazioni in un unico sistema d'informazione.

Le ISAG(G) sono costituite da 26 elementi divisi in 6 aree descrittive:

1. Area dell'identificazione.
2. Area delle informazioni sul contesto.
3. Area delle informazioni relative al contenuto ed alla struttura.
4. Area delle informazioni relative alle condizioni di accesso ed all'utilizzo.
5. Area delle informazioni relative alla documentazione collegata.

Nell'introduzione leggiamo ancora che l'applicazione delle regole può essere adattata alle esigenze specifiche:

The structure and content of the information in each of these elements should be formulated in accordance with applicable national rules. As general rules, these are intended to be broadly applicable to descriptions of archives regardless of the nature or extent of the unit of description.[...] However, the standard does not define output formats, or the ways in which these elements are presented, for example, in inventories, catalogues, lists, etc.

Non tutti gli elementi sono necessari; a livello internazionale sono definiti obbligatori i seguenti campi:

- a. Segnatura
- b. Denominazione o titolo
- c. Date della documentazione
- d. Consistenza dell'unità di descrizione
- e. Livello di descrizione.

La descrizione archivistica non può essere pienamente associata ed equiparata ad una descrizione bibliografica in quanto presenta alcune caratteristiche particolari. È un tipo di descrizione multilivellare con una precisa struttura gerarchica, di conseguenza è importante per la comprensione dei documenti che la descrizione di ogni unità sia collegata all'unità di livello superiore.

All'interno di ALEPH non si è sentita l'esigenza di creare legami gerarchici per la semplicità strutturale dell'archivio, anche se il formato UNIMARC lo avrebbe

permesso grazie al meccanismo dei legami gerarchici tra la collezione ed i volumi che ne fanno parte.

Per rendere esplicito il livello superiore di ogni elemento, nel processo di inserimento delle informazioni, nel campo della segnatura archivistica è stato inserito il codice dell'unità gerarchica più alta.

La descrizione a livello generale del fondo è riportata nel sito come introduzione all'archivio in modo da facilitare la ricerca e la comprensione dei documenti, mentre nel database sono state inserite le descrizioni limitatamente al livello dei fascicoli.

La tabella seguente è utilizzata per la conversione dei campi ISAD(G) in formato UNIMARC. Vengono qui riportati i campi effettivamente inseriti in ALEPH.

CAMPO ISAD(G)	CAMPO ALEPH	CAMPO UNIMARC
Segnatura archivistica od altri codici identificativi.	segnatura archivistica (abbreviazione del nome della serie, n. del faldone, n. del fascicolo [es. Did1.1])	LOC (proprietario)
Denominazione o titolo	Titolo o denominazione	200 - 1
Data della documentazione compresa nell'unità di descrizione	----- ---	
Livello di descrizione	Non è stato inserito perché è solo uno, cioè il fascicolo.	
Consistenza dell'unità di descrizione (quantità, volume, dimensione fisica)	Consistenza dell'unità e tipologia del materiale	215
Denominazione del soggetto produttore.	Non è stato inserito perché c'è un solo produttore	
Storia istituzionale/amministrativa, nota biografica.	Campo inserito nella descrizione a livello di fondo effettuata al di fuori di ALEPH	
Estremi cronologici dell'unità di descrizione.	Data della documentazione	210

Storia dei passaggi di responsabilità giuridica.	Campo inserito nella descrizione a livello di fondo effettuata al di fuori di ALEPH	
Modalità di acquisizione.	Campo inserito nella descrizione a livello di fondo effettuata al di fuori di ALEPH	
Illustrazione del contenuto/Abstract.	Abstract	ABS (proprietario)
Procedure e criteri di valutazione e scarto	Campo inserito nella descrizione a livello di fondo effettuata al di fuori di ALEPH	
Incrementi previsti.	-----	
Criteri di ordinamento.	Campo inserito nella descrizione a livello di fondo effettuata al di fuori di ALEPH	
Condizione giuridica.	Campo inserito nella descrizione a livello di fondo effettuata al di fuori di ALEPH	
Consultabilità.	Campo inserito nella descrizione a livello di fondo effettuata al di fuori di ALEPH	
Copyright/Condizioni che regolano la riproducibilità della documentazione.	Campo inserito nella descrizione a livello di fondo effettuata al di fuori di ALEPH	
Lingua della documentazione.	Lingua	101 - 0
Caratteristiche materiali.	Caratteristiche materiali	307
Strumenti di ricerca.	Campo inserito nella descrizione a livello di fondo effettuata al di fuori di ALEPH	
Localizzazione degli originali.	Campo inserito nella descrizione a livello di fondo effettuata al di fuori di ALEPH	
Esistenza di copie.	----- ---	
Unità di descrizione collegate, conservate nel medesimo istituto archivistico.	-----	

Materiale documentario complementare, conservato in altri istituti archivistici.	-----	
Bibliografia.	-----	
Area delle note.	Note	300

CAPITOLO 4: La gestione di immagini e metadati

Un'immagine digitale non ha una forma ed è visibile solo tramite il monitor di un computer o il televisore. Ciò significa che gli utenti non hanno riferimenti tangibili grazie ai quali ricordare o collocare un'immagine; l'unico modo per reperirla è usare i metadati ad essa associati¹.

Nell'era del computer "il metadato è considerato un collante tra molte risorse proprietarie come applicativi, tecnologie legate a internet ed ai database"². Gli strumenti per la gestione dei metadati, i *repository*, sono diventati un elemento cruciale in ambiente digitale. Secondo la definizione di Bernstein un repository "è un database condiviso di informazioni strutturate come software, documenti, mappe e altro"³.

Una delle caratteristiche che contraddistingue un repository è il tipo di metadati che gestisce; altre caratteristiche sono la presenza di un modello di informazione che descrive i tipi essenziali di metadati; di un linguaggio formale di specifica per i tool; di un linguaggio che supporta l'interoperabilità tra i tool dei diversi software e di un linguaggio standard di interrogazione per la ricerca dei metadati.

Un repository è caratterizzato poi dal common infrastructure support; la sua funzione è quella di gestire attività come l'integrazione dei tool, il supporto multi-piattaforma, la gestione degli eventi. Ad un livello inferiore troviamo il repository engine responsabile della creazione e della gestione degli objects, delle loro versioni e configurazioni. Solitamente il repository engine è progettato e sviluppato usando un modello object-oriented. Infine, un repository ha un data server che supporta l'inserimento, la cancellazione, la modifica ed il reperimento delle informazioni contenute nel repository.

¹ TECHNICAL ADVISORY SERVICE FOR IMAGES "Image Management Systems " in *Technical Advisory Service for image - Advice Paper*
<<http://www.tasi.ac.uk/advice/creating/pdf/ims2.pdf>>.

² ARUN SEN, "Metadata management: past, present and future", in *Decision Support System*, 37 (1) 2004: 151-173.

³ *ibidem*

In particolare, un Image Management System (IMS) si basa sulla struttura logica dei database allo scopo di organizzare, gestire e reperire le immagini ed i metadati ad esse associati. L'IMS immagazzina le immagini o i link alle immagini, insieme a tutti i metadati rilevanti per compiere facilmente e velocemente ricerche all'interno dell'archivio. Le funzioni basilari che un sistema di gestione delle immagini deve svolgere, secondo il TASI, sono⁴:

- memorizzare le immagini
- memorizzare i metadati relativi alle immagini
- fornire una facilitazione nella ricerca all'interno dei metadati delle immagini
- fornire un mezzo per controllare visivamente e scorrere le immagini

Come per tutti i sistemi di gestione dell'informazione elettronica, l'architettura e lo sviluppo di ogni singolo sistema dipende dai requisiti specifici dello stesso. Una soluzione ideale deve tenere conto delle finalità, dell'utenza e della funzione del sistema.

Nella scelta dell'IMS Peter Hirtle⁵, vicedirettore del Cornell Institute for Digital Collections, consiglia di valutare come il sistema sarà usato, stimandone:

- **Finalità.** È impossibile scegliere o costruire un IMS se non si sa quali sono gli obiettivi da raggiungere. Sarà un sistema ad accesso pubblico o si intende farne solo un uso interno? Sarà un database statico o dovrà essere in grado di supportare aggiornamenti e modifiche dei record? Gli utenti avranno bisogno di strumenti particolari per manipolare le immagini o lavoreranno con il materiale non elaborato del database?
- **Dimensione.** Bisogna innanzitutto stimare il numero e la dimensione delle risorse digitali. Un sistema che consta di decine di centinaia di immagini potrà essere meglio amministrato con un sistema di gestione relativamente semplice e poco costoso. Un sistema che contiene centinaia di migliaia di immagini ha bisogno di un software più complesso. Si deve considerare il

⁴ TECHNICAL ADVISORY SERVICE FOR IMAGES "Image Management Systems " in *Technical Advisory Service for image - Advice Paper*
<<http://www.tasi.ac.uk/advice/creating/pdf/ims2.pdf>>

⁵ PETER HIRTLE, „Image Management Systems and Web Delivery“ in ANNE KENNEY, OYA RIEGER, *op. cit.*, p. 120 – 122.

numero e la dimensione dei record, il numero degli utenti simultanei, l'utilizzo di un software dimensionato alle capacità dell'hardware e la performance dell'intero sistema. Si deve inoltre tenere conto della capacità della biblioteca di sviluppare e mantenere un database complesso.

- ***Complessità e volatilità dei dati.*** Un IMS può contenere dati relativamente semplici, per esempio solo il titolo e la collocazione dei file, oppure può diventare molto complesso con ampie informazioni riguardo alla descrizione, alla storia e all'uso del patrimonio rappresentato dai record nel database. Un database complesso può avere un migliaio o più campi collegabili da centinaia di tabelle. Inoltre, le risorse gestite da un IMS possono essere statiche, cioè dopo aver acquisito e identificato le immagini si può non avere bisogno di ulteriori aggiornamenti o modifiche dell'IMS, non essendo più necessario variare le descrizioni e le collocazioni dei file.
- ***Funzioni generiche dei database (autenticazione, requisiti di audit, sistema di monitoraggio e back-up).*** Nei sistemi complessi è spesso necessario sapere chi sta eseguendo certe funzioni e che cosa è autorizzato a fare. È perciò importante definire dei requisiti di sicurezza. Si dovrà chiedere agli utenti di registrarsi e di accedere al sistema con username e password. Si può inoltre decidere di creare classi di utenti abilitati a svolgere operazioni quali per esempio vedere le immagini in diverse dimensioni o visualizzare documenti ad accesso limitato.
- ***Infrastruttura tecnica disponibile (hardware, software e personale).*** La fase di creazione è il momento giusto per capire se ci sono possibilità di collaborazione con altre istituzioni. Nessuna risorsa è più importante dello staff tecnico. Molti IMS sono la combinazione di prodotti commerciali e di programmi personalizzati. Nessun sistema già pronto va incontro a tutte le esigenze di una biblioteca digitale, specialmente riguardo all'interfaccia utente, ai metadati ed alle strategie di migrazione. Uno staff tecnico deve integrare tutte le parti che compongono il sistema e mantenerle funzionanti. Un investimento sulla formazione di esperti all'inizio del progetto può risultarne in seguito la salvezza.
- ***Costi.*** Bisognerebbe prendere in considerazione i costi complessivi del sistema quando si sceglie un IMS, ma è normalmente impossibile valutare i

benefici che si possono trarre dall'implementazione di un sistema piuttosto che di un altro. Non c'è una letteratura che affermi che i benefici derivanti da un buon sistema di gestione delle immagini non superino le spese di acquisizione e mantenimento del sistema. Ogni istituzione deve determinare quanto denaro può spendere per migliorare l'accesso.

Si devono inoltre considerare le caratteristiche generali dell'architettura di sistema e come possono contribuire ad un uso efficiente dello stesso⁶.

- ***Flessibilità.*** Considerando la rapida evoluzione degli applicativi per le immagini, l'architettura dell'IMS sarà probabilmente soggetta a cambiamenti. È perciò imperativo che il sistema sia flessibile in modo che possa essere facilmente implementato.
- ***Database singoli o multipli.*** La via più sicura per assicurarsi un'architettura flessibile è sviluppare diversi database, ognuno con un compito particolare. In questo modo ogni sistema può essere ottimizzato per il suo particolare uso; la separazione delle componenti può inoltre fornire maggiore sicurezza e l'eccessiva richiesta su un sistema non influisce sulla performance degli altri.
- ***Fedeltà agli standard del data system, della strutturazione e del contenuto.*** L'uso di standard nel data system, nella strutturazione dei dati e nel contenuto assicurano flessibilità in quanto rendono più semplice lo scambio di informazioni tra sistemi e strumenti di supporto nella migrazione dei dati.
- ***Durata della vita dei dati prevista e possibilità della migrazione.*** Ci sono molti modi per assicurarsi che i metadati all'interno di un IMS attuale continueranno ad essere validi e fruibili in futuro, per esempio progettando il sistema con un chiaro percorso di migrazione per i dati. La creazione di dati non strutturati e non controllati non potrà mai avere un futuro anche nel più aperto sistema immaginabile. La fedeltà agli standard di metadati o almeno l'adozione di dati strutturati rimangono la garanzia più sicura della futura vitalità dei dati.

⁶ PETER HIRTLE, „Image Management Systems and Web Delivery“ in ANNE KENNEY, OYA RIEGER, *op. cit.*, p. 123 – 125.

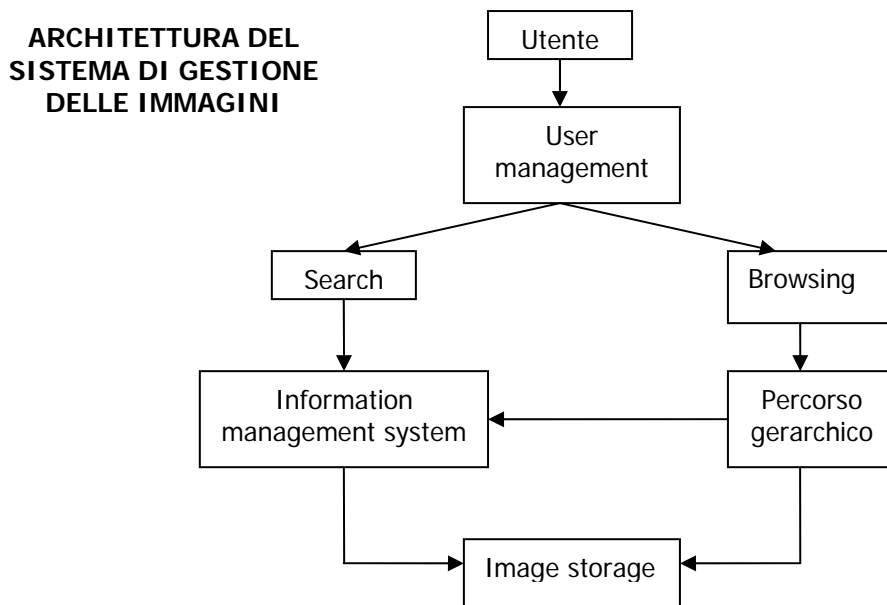
Per molte applicazioni le biblioteche si affidano a programmi client/server, tra queste si collocano i cataloghi in linea. Secondo Peter Hirtle⁷ i cataloghi esistenti presentano molti vantaggi se usati come sistemi di gestione di immagini ed informazioni collegate in quanto "rappresentano un enorme investimento nei metadati preesistente. Diventa sensato provare a sfruttare questa risorsa piuttosto che crearne una nuova". Inoltre, continua Hirtle, "i nuovi cataloghi delle biblioteche permettono di collegare facilmente i metadati contenuti nel catalogo e le immagini digitali", ed infine "molti cataloghi sono mantenuti da uno staff tecnico altamente qualificato", ciò significa che il progetto può trarre vantaggio dalla presenza di uno staff esperto. Ci sono anche degli svantaggi in questo tipo di scelta, segnala sempre Hirtle: molti cataloghi si basano sulla struttura del MARC, formato che può non adattarsi facilmente ad alcuni metadati; inoltre la mescolanza di record di diverso livello può portare a problemi associati alla granularità dell'informazione.

Al momento le alternative più accreditate sono soluzioni basate sullo standard SGML o sulla XML Document Type Definition (DTD). La fortuna di questi due linguaggi, basati sull'uso di marcatori, è la loro adattabilità alla descrizione dei diversi tipi di materiale.

Per il progetto Benussi è stato pensato un sistema basato su un'architettura client/server e su un'elaborazione distribuita che permettesse una maggiore flessibilità.

⁷ PETER HIRTLE, „Image Management Systems and Web Delivery“ in ANNE KENNEY, OYA RIEGER, *op. cit.*, p. 129.

LA STRUTTURA DEL DATABASE



Ogni componente dell'architettura è stata sviluppata e viene gestita separatamente, in modo che sia possibile l'implementazione di singoli aspetti del sistema senza dover modificare l'intera struttura.

Lo user management si svolge a due livelli. Ad un primo livello vengono analizzate tutte le operazioni svolte dall'utente all'interno del sito attraverso un log analyzer installato sul server web della biblioteca. Ad un secondo livello, all'interno di ALEPH, il sistema è in grado di produrre delle statistiche d'uso e di tenere memoria delle ricerche eseguite dall'utenza; questo dovrebbe aiutare a capire le strategie di ricerca più frequentemente utilizzate.

Nell'implementazione del sistema di gestione delle informazioni si è ritenuto prioritario raggiungere i seguenti obiettivi:

- La realizzazione di un database che contenga le informazioni per localizzare i documenti digitali.
- L'inserimento di informazioni necessarie per migliorare il reperimento delle immagini.
- La creazione di un inventario scevro da errori e standardizzato
- L'organizzazione delle informazioni nel rispetto della struttura organico-funzionale dell'archivio.

- L'accesso diretto alle informazioni attraverso la ricerca per denominazione, data, luogo e parola chiave.

La scelta è stata quella di sfruttare il software che già gestisce il patrimonio della Biblioteca, ALEPH, per varie motivazioni:

- Il sistema riesce a soddisfare le finalità di Information Management System, cioè di permettere un uso pubblico del sistema, l'implementazione e la modifica dei dati ed un accesso semplice alle immagini.
- Le esigue dimensioni dell'archivio, che richiede il salvataggio di circa 300 record di descrizione ed il reperimento di circa 12.000 immagini, hanno reso superflua la creazione o l'acquisto di uno strumento ad hoc.
- La struttura dei dati, non molto complessa, è stata agevolmente adattata alle potenzialità del sistema.
- La struttura aperta del sistema permette una forte parametrizzazione e personalizzazione di ogni aspetto e funzione da esso svolta, assicurando una grande flessibilità.
- L'uso di uno strumento già in possesso della biblioteca ha notevolmente ridimensionato i costi ed ha richiesto solo un impegno lavorativo che sarebbe comunque stato necessario per l'implementazione di un qualunque sistema.
- Questo tipo di scelta non ha inoltre comportato l'acquisto di un nuovo hardware, in quanto la biblioteca si avvale dei server del Consorzio Interuniversitario Lombardo, ed ha permesso di affidarsi allo staff tecnico della biblioteca senza ulteriore bisogno di formazione del personale.
- Diverse unità sono preposte ai distinti compiti che il sistema svolge e l'uso di differenti library permette di mantenere separato il materiale librario da quello d'archivio.
- La scelta consente la fedeltà agli standard ed ai protocolli di data system, in quanto il sistema si basa sul linguaggio di interrogazione SQL, quello attualmente più usato, e consente l'integrazione con altri database grazie ad un modulo di interrogazione aggiuntivo basato sullo standard Z39.50. Permette inoltre aderenza agli standard per la strutturazione dei dati grazie all'uso dell'UNIMARC, largamente usato nella comunità biblioteconomica.

- All'interno della biblioteca è già stata verificata la funzionalità e la buona riuscita delle migrazioni dei record bibliografici e degli altri dati da un altro sistema ad ALEPH e da una versione del software ad una più recente.
- L'esistenza di altre raccolte digitali, in particolare negli Stati Uniti⁸, gestite con lo stesso software e di biblioteche italiane che lo usano per la descrizione di fondi archivistici (l'Università degli studi di Siena ed il Politecnico di Torino) ha assicurato sul buon esito di questa scelta.

Il software ALEPH è un sistema integrato di componenti hardware e software per la gestione dei dati bibliografici, implementato dalla ATLANTIS, società che gestisce applicativi per i client attraverso il sistema APIs (Application Program Interfaces).

Nella sua documentazione ATLANTIS fornisce questa descrizione del sistema:

ALEPH's architecture is based on a multitier, client/server model. Client/Server communication is based on a stateless (selfcontained) transaction model, nonetheless, ALEPH Application Servers keep continuous connections (with time-out) to the database, to ensure high performance. ALEPH features a flexible database design.⁹

Il sistema ALEPH è composto di cinque unità tra loro strettamente collegate: un'unità di Authority Control, un'unità bibliografica, un'unità per la gestione del posseduto, un'unità amministrativa e un'unità di amministrazione a livello generale di sistema. Il modello di database supporta un ampio spettro di configurazioni ed implementazioni, compresa l'installazione indipendente delle unità su diversi host, e gioca un ruolo cruciale nella creazione di un sistema multilivello.

⁸ Il Jewish Theological Seminary di New York ha creato una collezione digitale di testi sacri consultabile all'indirizzo <<http://www.jtsa.edu/library/digitalcollections.shtml>>.

⁹ Documentazione ATLANTIS.

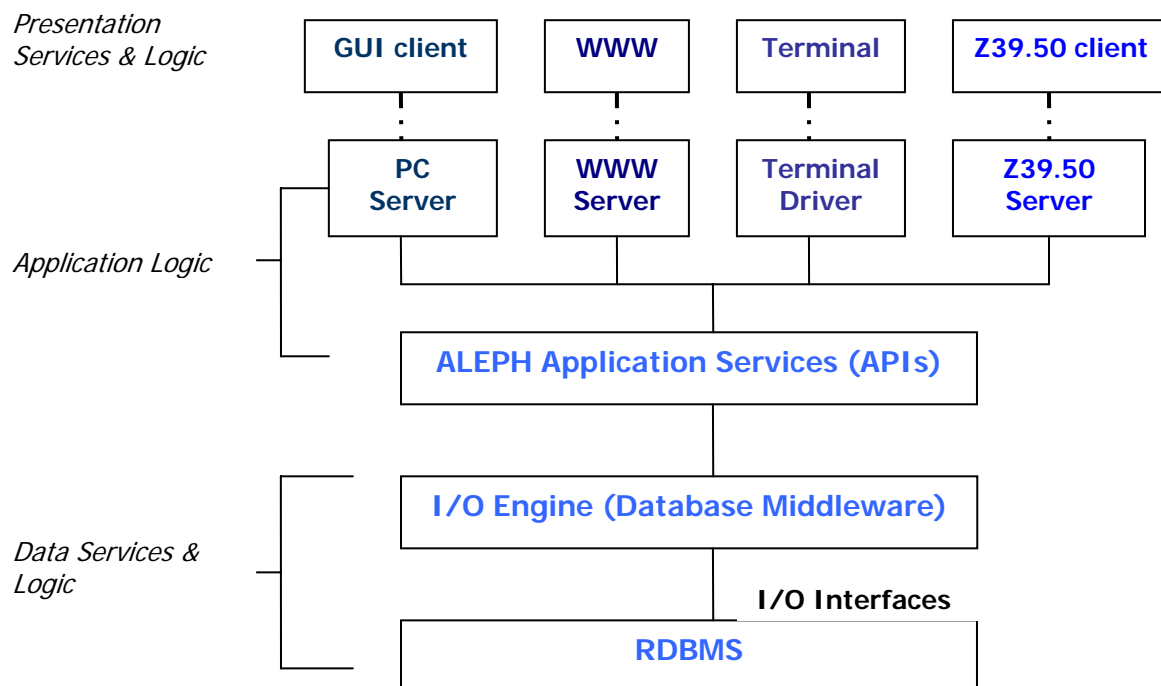


Fig. 2: Architettura del Database¹⁰

Il server ALEPH, come si può vedere dalla figura 2, è suddiviso in tre livelli:

- **Presentation Services & Logic**

- Fornisce l'interfaccia con cui l'utente interagisce.

- **Application Logic**

- *Livello degli Application Servers* – Questo livello è composto da server per ogni interfaccia. Ogni application server riceve una domanda da un client, la traduce in un formato omogeneo e la indirizza all'Application Service object (API). Una volta tradotte, tutte le domande hanno lo stesso formato indipendentemente dalla loro origine (Z39.50 client, ALEPH client, WWW browser o terminale).
- *Livello degli Application Services (APIs)* – È il cuore di ALEPH. Questo livello è composto da un insieme di Application Program Interface, che provvedono alle operazioni per i vari client.

- **Data Services & Logic**

- *ALEPH Database Middleware (o I/O Engine)* – È lo strato di gestione del database ad alto livello. I data services vengono forniti agli application services objects da un server logico che contiene un gruppo di objects che si

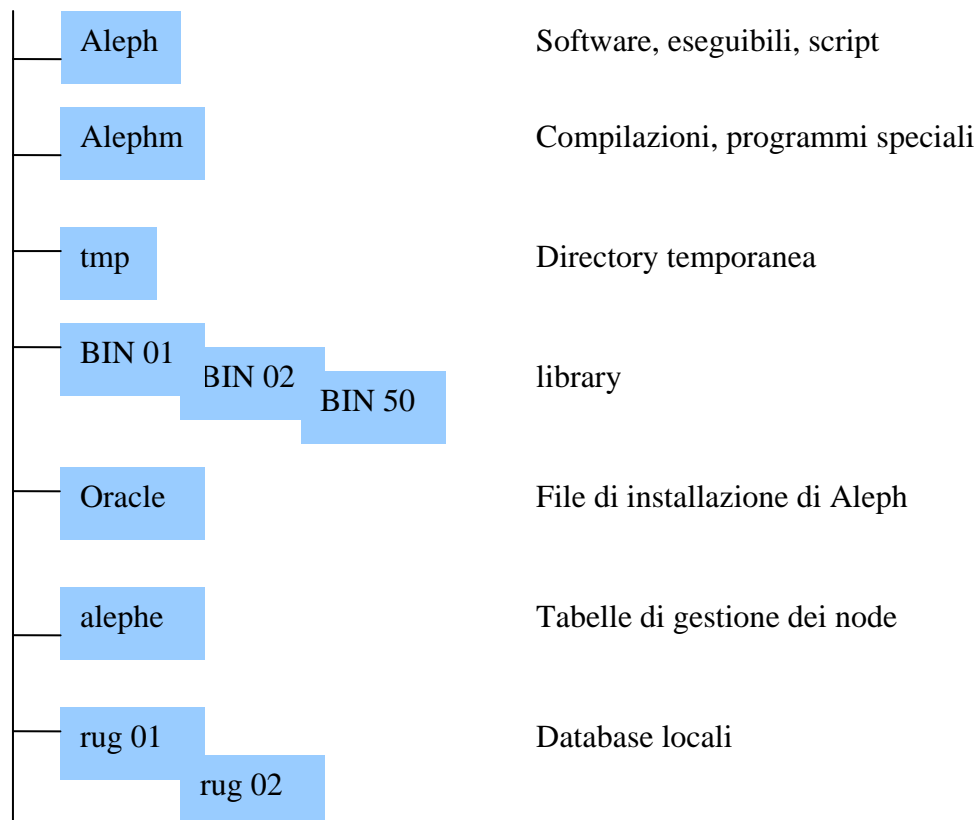
¹⁰ Documentazione ATLANTIS

pongono come intermediari tra l'applicazione ed il database. Il motore I/O traduce la richiesta di un applicativo in una sequenza di comandi per il database e fornisce un'implementazione di SQL adattabile alla natura dei dati catalografici, testuali e non formattati. Il motore I/O sfrutta anche la conoscenza del sistema delle caratteristiche speciali dei DBMS, al fine di ottimizzare l'aggiornamento ed il reperimento delle informazioni. In questa partizione logica ALEPH associa la sua esperienza ed il suo know-how alle particolari strutture dei dati ed ai particolari formati delle biblioteche. L'intermediazione del motore I/O tra le applicazioni ed il DBMS assicura la massima flessibilità della struttura logica e fisica del DBMS.

- *ALEPH Database* – Il database di ALEPH è basato su un RDBMS Oracle. Per la natura testuale e non formattata dei cataloghi bibliografici, ALEPH è implementato con il linguaggio SQL per assicurare più alte prestazioni.

Un aleph root è composto di tre tipi di directory:

- | | |
|--------------------|---------------------------------------|
| 1. ALEPH software | alephm, aleph, tmp, usm01, usm10, ... |
| 2. Node management | alephe |
| 3. Libraries | bin01, bin02, ... |



La directory **alephm** (alias "am") contiene i programmi e le procedure di installazione, sviluppo e mantenimento del software; la directory **aleph** contiene, invece, i moduli run time di **ALEPH**. **Alephe** include, infine, le tabelle di gestione dei Node ed i file di parametrizzazione riguardanti il server. Ogni library di ALEPH ha una diversa directory di origine che è parte di questo albero e che contiene informazioni rilevanti per gestire la library stessa.

ALEPH500, come detto in precedenza, è basato su un RDBMS Oracle 9i, che consiste di unità logiche chiamate tablespaces. Ogni tablespace è costituita da uno o più datafile che possono essere salvati su uno o più dischi.

Tablespace Name	Usage	Physical File
System	Oracle system tables	/exlibris/oradata/aleph0/system01.dbf
Temp	Temporary space (for sorting, index creation, etc...)	/exlibris2/oradata/aleph0/temp01.dbf
ts0	ALEPH tables	/exlibris2/oradata/aleph0/ts0_0.dbf
ts1	ALEPH indexes	/exlibris/oradata/aleph0/ts1_0.dbf

All'interno di Oracle per ogni library viene implementato un database, ognuno dei quali possiede un suo set di tabelle di configurazione e un suo set di tabelle all'interno delle quali si trovano i dati relativi alla library. Questa suddivisione in library serve ad organizzare i dati in modo logico, così che i dati bibliografici e di Authority control possono essere condivisi con altre biblioteche, mentre i dati amministrativi rimangono separati.

Per il Fondo Benussi non si è ritenuto necessario mantenere la divisione tra dati bibliografici, amministrativi, sulle copie e di authority control, pertanto tutte le informazioni riguardanti l'archivio risiedono in un'unica library (BIN02).

Per l'accesso diretto all'informazione sono stati creati quattro tipi di indici:

- denominazione.
- lingua del documento
- data di creazione
- parola-chiave.

Le parole-chiave sono state assegnate ai record senza il supporto di un vocabolario controllato, ma sono stati adottati i termini maggiormente accreditati in campo psicologico.

Gli indici di Oracle sono Oracle object (B-tree) che contengono puntatori a una riga specifica di una tabella. Ogni indice è mappato da una tablespace. Per ogni indice Oracle assegna uno spazio iniziale ed uno spazio esteso, secondo la specificazione nel comando CREATE INDEX, controllati da una speciale configurazione chiamata **file_list**.

L'uso di un Information Management System, rispetto agli inventari cartacei, ha il vantaggio di creare indici in riferimento a soggetti, persone, luoghi complementari all'inventario e di agevolare la ricerca per parole-chiave anche se questi indici, per la loro natura, non possono essere conformi a norme standardizzate.

FILE NAMING SYSTEM

"Oltre ad essere descritte, le risorse digitali hanno bisogno di essere identificate con un nome univoco e persistente"¹¹, dice la Tammaro.

La scelta del file naming è un argomento che deve essere valutato all'inizio del progetto di digitalizzazione e che deve essere incluso nelle specifiche tecniche del progetto. La documentazione del TASI è tra le poche ad offrire delle linee guida anche nell'assegnazione dei nomi ai file di immagine¹².

Attente valutazioni devono essere compiute ogniqualvolta si decide come nominare o numerare i file in una biblioteca digitale. Un sistema di file naming adeguato non solo assicurerà una certa coerenza, ma potrà anche essere parte integrante del processo di ricerca.

Affinché il nome di un file sia compatibile tra le varie piattaforme, esso dovrebbe essere assegnato in base alla convenzione 8.3, ideata per la piattaforma DOS, secondo cui i nomi dei file sono limitati ad otto caratteri, seguiti da un'estensione di tre caratteri (per esempio, topolino.jpg). Il rispetto

¹¹ ALBERTO SALARELLI, ANNAMARIA TAMMARO, *op. cit.*, p. 183 – 184.

¹² TECHNICAL ADVISORY SERVICE FOR IMAGES "Image Management Systems " in *Technical Advisory Service for image - Advice Paper*
<<http://www.tasi.ac.uk/advice/creating/pdf/ims2.pdf>>.

di questa convenzione diventa particolarmente importante quando si masterizzano i file su CD-R o DVD-R, dato che l'ISO 9660 file naming standard¹³ usa lo standard 8.3 per assicurare l'interoperabilità tra tutte le piattaforme.

I nomi dei file si suddividono in "descrittivi" e "non descrittivi". Il nome di un file è descrittivo quando è costituito da parole reali, abbreviazioni o numeri che abbiano qualche relazione con il contenuto, mentre non è descrittivo quando consiste in un numero, o in una combinazione di numeri e lettere, che non hanno significato per chi li legge. Un sistema di file naming descrittivo tende ad adattarsi meglio alle raccolte più piccole dove è facile scorrere visivamente l'insieme delle cartelle e dei file nominati in modo descrittivo.

Lo svantaggio di questo modo di procedere è che i nomi dei file devono essere assegnati manualmente o con procedure solo parzialmente automatizzate, aumentando così la probabilità di errori, o che lo stesso nome venga assegnato a più di un file.

L'assegnazione di nomi non descrittivi ai file, invece, si adatta meglio alle raccolte più grandi, o a quelle che usano più sofisticati sistemi di gestione delle immagini (IMS), che normalmente generano in modo automatico nomi univoci. Una volta completata la digitalizzazione, ogni file ha un numero significativo solo per la macchina, ma con la garanzia che sia univoco. Lo svantaggio di questo sistema è che non risulterà molto utile, dal punto di vista della ricerca e del recupero delle immagini, se non è associato ad un Image Management System che connetta i numeri con i loro metadati indicizzati. Per questo motivo i nomi dei file non descrittivi si adattano meglio alle raccolte più ampie, in quanto una raccolta, troppo vasta per essere navigata, fa di solito affidamento, per localizzare i file, su un database che rende inutile l'attribuzione di nomi analitici. L'uso del 8.3 naming standard permette di creare 999,999,999 filename unici, e, se qualche numero è sostituito con lettere, questo sistema può coprire praticamente qualsiasi dimensione della collezione. A differenza dei nomi descrittivi, che possono aiutare ad identificare i file di immagine

¹³ Il filesystem ISO 9660 è lo standard usato per i CD-ROM; è particolare a causa della struttura stessa dei CD: i dati vengono memorizzati in settori su un'unica traccia a spirale che parte dalla zona centrale e si espande verso l'esterno. Il filesystem ISO 9660 è predisposto per gestire file il cui nome è organizzato nello stesso modo in cui faceva il Dos: 8.3, ovvero otto caratteri al massimo, seguiti da un punto e da un'estensione di un massimo di tre caratteri.

indipendentemente dal loro database, un sistema di numerazione sequenziale fa sì che i file siano identificati solo tramite i metadati all'interno del database. La protezione dell'Image Management System diventa, in questo caso, un punto focale nella strategia di preservazione del digitale alla pari della protezione delle immagini stesse: senza l'IMS le immagini, nominate in modo casuale, risulterebbero irreperibili.

I file di immagine creati per il reperimento sul web devono seguire alcune ulteriori convenzioni. Prima di tutto, è meglio usare solo caratteri minuscoli in considerazione del fatto che alcuni sistemi operativi, come Unix, sono case-sensitive. In secondo luogo, viene raccomandato di evitare completamente caratteri speciali nei nomi dei file perché alcuni simboli come \ / : * ? " < > | rivestono funzioni specifiche all'interno dei diversi sistemi operativi. Possono invece essere impiegati il trattino e l'underscore per sostituire gli spazi visto che i nomi dei file destinati al web non ne possono includere.

Nel caso di un archivio è abbastanza semplice riuscire a creare un sistema di file naming descrittivo seguendo la struttura gerarchica del fondo stesso. Si deve comunque considerare il rischio che la combinazione di lettere e numeri non abbia un significato così chiaro per gli utenti e che quindi il senso del sistema di assegnazione di nomi significativi ai file risulti un lavoro inutile dal loro punto di vista.

Inizialmente i file TIFF ottenuti dal processo di digitalizzazione hanno ricevuto una numerazione sequenziale non significativa. Per poter comunque individuare facilmente le immagini, in mancanza di un IMS, è stata creata una struttura di cartelle e sottocartelle che rispecchiava quella dell'archivio. Quasi subito si è reso evidente che purtroppo la numerazione non era univoca e che per i fascicoli di poche pagine questa non era neanche stata attribuita, come richiesto, dalla società che ha svolto il lavoro di digitalizzazione. Durante il lavoro di esegesi sui documenti, si è inoltre presentata la necessità di dover spostare alcune immagini da un fascicolo ad un altro. La numerazione, così come era stata attribuita, rischiava di creare confusione per cui è stato necessario attribuire una nuova numerazione. È sembrato, a questo punto, opportuno attribuire ai file dei nomi significativi in base ad un sistema alfanumerico:

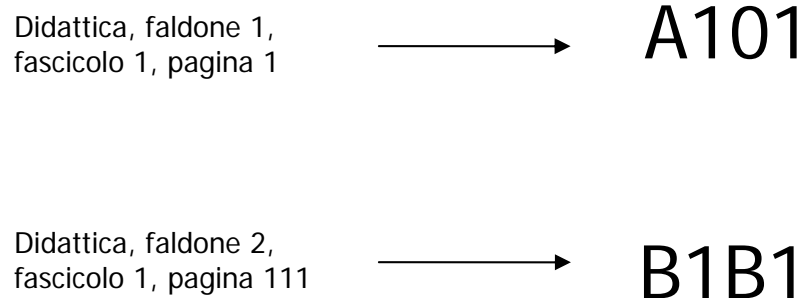


Fig. 5: Esempi di Filename

Come detto sopra, una notazione può essere composta da un numero massimo di otto caratteri alfanumerici. Si è però stabilito di utilizzare solo i primi quattro e di destinare eventualmente i quattro spazi successivi alla vecchia notazione nel caso in cui ne venga attribuita una nuova a documenti spostati in un momento successivo, in modo da mantenere una specie di percorso d'origine del documento stesso.

La lettera iniziale indica il faldone in cui il documento è conservato. Ai faldoni sono state associate le lettere dell'alfabeto visto che erano sufficienti. Convenzionalmente si è deciso che le prime dieci lettere (dalla lettera A alla lettera L) indicano i dieci faldoni della Didattica nel loro ordine, l'undicesima lettera (M) si riferisce al faldone delle Pubblicazione, ed infine la dodicesima lettera (N) contraddistingue il faldone contenente la serie del Carteggio.

La seconda posizione indica il numero del fascicolo. All'interno dei faldoni si trovano al massimo 60 fascicoli e, per poterli identificare tutti, vengono usati prima i numeri da 0 a 9, poi le lettere maiuscole e infine le lettere minuscole. La terza e la quarta posizione identificano il numero di pagina, la terza posizione rappresenta con caratteri numerici le decine e con i caratteri alfabetici le centinaia, mentre nella quarta posizione si trovano solo caratteri numerici da 0 a 9 che rappresentano le unità.

Analogamente si è proceduto nell'assegnazione dei filename alle immagini in formato JPEG create per la visualizzazione sul web.

FALDONE	FASCICOLO	PAGINA(decine)	PAGINA (unità)
Did1 : A	0	0	0
Did2 : B	1	1	1
Did3 : C	2	2	2
Did4 : D	3	3	3
Did5 : E	4
Did6 : F	9	9	5
Did7 : G	A	A	6
Did8 : H	B	B	7
Did9 : I	C	C	8
Did10 : J	9
Cart : K	Z	Z	-----
Pubb : L	a	a	
	b	b	
	c	c	
	
	Z	Z	

Fig. 6: Schema completo del Filename

CAPITOLO 5: L'interfaccia e l'accesso.

PRESENTAZIONE: IL SITO

In senso generale si può definire interfaccia qualunque strumento ci aiuti ad interagire col mondo intorno a noi in modi il più possibile adatti alla nostra conformazione fisica e sensoriale. Ad esempio Fabio Ciotti e Gino Roncaglia¹ definiscono anche la forchetta ed il coltello "interfacce che ci mettono in relazione con il cibo" e, restringendo la definizione al solo ambito informatico, danno una definizione di interfaccia come l'insieme dei dispositivi hardware e software che permettono di interagire con una macchina o un programma in maniera il più possibile semplice ed intuitiva.

L'interfaccia web determina, pertanto, in modo rilevante il successo della biblioteca digitale, in quanto strumento di intermediazione, realtà con cui l'utente entra in contatto e con cui si confronta. I requisiti di semplicità ed intuitività a cui l'interfaccia deve rispondere dipendono largamente da una componente sociale e culturale.

Gli studi ripetutamente condotti nell'ambito del progetto DLI² sull'impatto che i progetti di Biblioteca Digitale avevano sull'utenza hanno dimostrato che tra le cause di insuccesso c'è la difficoltà da parte dell'utente di consultare tale mezzo. La progettazione dell'interfaccia deve innanzitutto rispondere ad un'esigenza di semplicità, di linearità e di immediatezza, senza però dimenticare che anche la veste grafica ha una sua importanza.

Nell'ideazione e costruzione delle pagine web un notevole peso rivestono gli studi fatti per migliorare l'interfacciamento tra l'uomo e il computer.

L'USABILITÀ

Alla base di questo concetto sta l'idea che si deve favorire la ricerca di informazioni e la conoscenza da parte dell'utente, quindi ogni progetto deve essere incentrato sull'utente.

¹ FABIO CIOTTI, GINO RONCAGLIA, *op. cit.*

² BETH SANDORE, "What Users Want from Digital Image Collections" in ANNE KENNEY, OYA RIEGER, *op. cit.*

Intorno ai primi anni '90, già prima dell'avvento del web, Jacob Nielsen, in collaborazione con Rolf Molich, elaborò dieci regole euristiche, cioè dieci regole empiriche³ nella creazione dei software.

1. **Visibilità dello stato del sistema.** L'utente dovrebbe sempre sapere qual è lo stato del sito in qualsiasi momento.
2. **Associazione del sistema con il mondo reale.** Il sistema deve usare il linguaggio naturale dell'utente, non la terminologia tecnica o del sistema stesso e deve osservare le convenzioni del mondo reale, fornendo mapping naturali.
3. **Controllo e libertà dell'utente.** L'utente dovrebbe sempre essere in grado di annullare e ripristinare qualsiasi azione stia compiendo senza rischiare di bloccare il browser.
4. **Coerenza e standard.** L'uso di standard e di regole precise nella progettazione dei software permette all'utente di capire più facilmente l'uso dei programmi stessi.
5. **Prevenzione degli errori.**
6. **Riconoscimento piuttosto che memorizzazione.** L'uomo raramente si trova nella necessità di dover ricordare tutte le caratteristiche di un oggetto a memoria. Le informazioni che gli servono possono risiedere nella sua mente (informazioni memorizzate) e nel mondo (etichette, informazioni "ovvie"). Il successo di un'interfaccia sta nel trovare una soluzione di equilibrio tra queste due opzioni delle informazioni necessarie per operare.
7. **Flessibilità ed efficacia d'impiego.**
8. **Estetica e progettazione minimalista.** Aggiungere elementi ad una pagina non significa per forza migliorarla, anzi in molti casi implica un peggioramento.
9. **Aiutare gli utenti a riconoscere gli errori, a diagnosticarli e a risolverli.**

³ Le dieci regole sono tratte dall'articolo di Jacob Nielsen "Ten usability Heuristics" <http://www.useit.com/papers/heuristics/heuristic_list.html>, L'elenco qui riportato è invece tratto dal volume MARK PEARROW, *Web usability: navigare, creare, gestire la rete*, Milano, Futura, 2001.

10. **Aiuto e documentazione.** Un qualche tipo di aiuto on-line è sempre necessario, anche sotto forma di una piccola lista di FAQ.

Negli ultimi anni Nielsen ha compiuto degli studi specifici sull'usabilità dei siti web⁴, aggiungendo a quelle già fissate nuove regole.

- **Si devono mantenere bassi i tempi di download e di risposta.** Tempi rapidi di risposta nel caricamento delle pagine sono il criterio più importante nella progettazione di un sito web. È stato calcolato che, dopo un secondo, il flusso dei pensieri di una persona si interrompe e che dopo circa 10 secondi il navigatore perda concentrazione nei confronti del sito. In 10 secondi si possono caricare tra i 34 Kb (con un modem) ed i 2 Mb (con una linea dedicata). La dimensione ideale di una pagina quindi non dovrebbe superare i 34 Kb, in quanto la maggior parte degli utenti si collega via modem. Nel caso specifico del progetto Benussi, considerando come target una comunità scientifica che dovrebbe usare un collegamento con una linea veloce, si possono costruire pagine con dimensioni fino a circa 100 Kb.
- **Le informazioni importanti devono stare in alto.** Se l'utente può iniziare a visualizzare una prima schermata di informazioni rilevanti, le pagine potranno anche avere un peso maggiore perché non sarà determinante se le immagini arrivano in un momento successivo. Affinché l'inizio della pagina sia significativo dovranno esservi collocate poche immagini, dovrà avere invece maggiore spazio il testo. Risulta inoltre utile l'uso dell'attributo <ALT> che permette di comprendere il contenuto delle immagini prima che queste vengano visualizzate. È infine meglio evitare di inserire tabelle troppo complesse all'interno delle pagine, perché il browser impiegherebbe più tempo per caricarle.
- **La presentazione deve essere indipendente dallo spazio.** Il grande successo di Internet deriva dal fatto che è uno strumento multiplatforma, vale a dire che un utente può accedere alla rete da un PC, da un portatile, da un palmare, dal suo cellulare. Anche gli schermi stessi dei PC possono avere dimensioni molto variabili. È quindi importante consentire a tutti di visualizzare il contenuto della pagina web nel modo corretto. Nielsen

⁴ JACOB NIELSEN, *Web usability*, Milano, Apogeo, 2000.

propone come standard, per la creazione delle pagine web, una risoluzione di 1024x768 pixel, pari ad uno schermo di diciassette pollici, in quanto secondo lui è quella più diffusa. In realtà in Italia schermi di tale dimensione non sono così diffusi come evidentemente lo sono in America. Di solito i siti italiani sono ottimizzati per una risoluzione di 800x600.

- **L'accessibilità del sito deve essere possibile da un qualsiasi punto della sua struttura.** Nel momento in cui le pagine web vengono costruite non è possibile determinare quali saranno le modalità di navigazione seguite dall'utente; potrà accadere che l'utente scelga percorsi di navigazione che al creatore non verrebbero mai in mente. Di conseguenza, poiché è anche importante che l'utente capisca in ogni momento in che punto del sito si trovi, il web designer dovrà cercare di creare una struttura del sito il più chiara possibile per l'utente.
- **Punto di accesso preferenziale e biglietto da visita per il sito è l'homepage.** L'homepage, che costituisce anche la directory dei contenuti principali del sito, sarà nella navigazione il punto di partenza e potrà essere diversificata dalle altre pagine pur condividendone lo stile.
- **Il nome del sito deve essere ripetuto in ogni pagina.** Esso permetterà in questo modo all'utente di capire in quale sito è entrato se non accede dall'homepage; tale scritta può essere riportata anche con un carattere non molto grande, ma deve essere in una zona ben visibile.
- **In ogni pagina deve essere presente il link all'homepage,** collocato sempre nello stesso punto, meglio se in alto a sinistra.
- **Nella costruzione delle pagine è opportuno evitare qualsiasi tipo di metafora,** perché essa rischia di essere fraintesa o non capita dall'utente. Per lo stesso motivo è preferibile anche evitare termini nuovi o poco conosciuti, facendo riferimento a quelli standard.
- **I collegamenti devono essere associati ad espressioni verbali,** preferibilmente non superiori alle 4 parole, che spieghino dove il collegamento conduce, infatti l'uso di immagini o di espressioni come "clicca qui" risultano meno facilmente comprensibili.

I link sono un elemento fondamentale dell'ipertesto e possono essere di tre tipologie:

- a. Topici, quando conducono ad ulteriori informazioni sull'argomento trattato;
- b. Strutturali, quando connettono i diversi livelli della struttura del sito;
- c. Associativi, quando indicano pagine dal contenuto simile.

Per diminuire il disorientamento dell'utente durante la navigazione può risultare utile attribuire dei titoli tramite l'uso dell'attributo title nel tag ``, non cambiare il colore di default dei link e dare brevi notizie sul contenuto dei siti quando si crea una lista di link esterni.

- **L'interfaccia deve sempre rispondere a tre domande fondamentali del navigatore:** "Dove mi trovo? Dove sono stato? Dove posso andare?" In questo modo l'utente conosce sempre la sua posizione in relazione al web ed alla struttura del sito. È molto utile per questo motivo che la presenza della struttura del sito rifletta il modello che si intende dare all'utente e che rappresenti le operazioni che gli utenti possono svolgere. Una struttura lineare può risultare poco chiara; è quindi meglio affidarsi ad una struttura gerarchica che indichi tutti i livelli direttamente sull'homepage e ad ogni livello tutte le alternative possibili.
- **Bisogna prestare attenzione alla scelta dei colori:**
 - a. Ci deve essere un forte contrasto tra i colori dello sfondo e quelli del testo, in positivo piuttosto che in negativo in modo che la vista non venga affaticata nel processo di lettura;
 - b. è da evitare l'abbinamento verde – rosso e le sue sfumature;
 - c. è preferibile per lo sfondo l'uso di colori pieni o di motivi molto delicati senza ricorrere ad immagini troppo complesse.
- **Non si deve usare una sola tecnica nel ritoccare le dimensioni delle immagini.** Solitamente i programmi di grafica consentono di ritoccare le dimensioni con la funzione di riduzione, in questo modo però l'immagine perde definizione e dettaglio. La tecnica del "cropping" permette invece di conservare i dettagli interni, ma ha lo svantaggio di far perdere il contesto dell'immagine originale. La soluzione migliore diventa quella di abbinare le due tecniche con una riduzione di immagine e con un'accentuazione della rilevanza.

- **È necessario evitare l'impiego ingiustificato dell'animazione.** Il suo uso deve essere ridotto al minimo, anche se essa può essere utile per determinati scopi:
 - a. Dà senso di continuità alle transizioni. Quando un passaggio può avere due o più stati possibili diventa più intuitiva una rappresentazione grafica che fa sì che l'utente usi, per compiere una deduzione, il sistema percettivo piuttosto che quello conoscitivo.
 - b. Serve a mostrare la direzione delle transizioni: una freccia per esempio aiuta maggiormente ad avere il senso della direzione che sta prendendo la navigazione.
 - c. Riesce ad illustrare bene l'evoluzione nel corso del tempo, in quanto l'animazione stessa è una rappresentazione visiva che cambia nel tempo.
 - d. È utile nel mostrare un insieme di più dati in una stessa porzione di schermo: l'esempio più classico è quello della mappa navigabile, sensibile al passaggio del mouse.
 - e. Arricchisce presentazioni grafiche che sarebbero più difficili da visualizzare con una grafica statica.
 - f. Serve a rappresentare strutture tridimensionali
 - g. Attira l'attenzione dell'utente su qualche aspetto significativo.

È importante che l'animazione non venga collocata a lato del testo in quanto l'immagine diventa fonte di distrazione per l'occhio umano. Alcuni utenti sostengono che le animazioni sono segno dell'accuratezza grafica del sito, ma nella maggior parte dei casi esse infastidiscono l'utente, soprattutto quando si tratta di testi scorrevoli. Le scritte lampeggianti invece vengono ormai collegate mentalmente ai banner pubblicitari, quindi ne è sconsigliabile l'uso.

Mark Pearrow, nel suo libro sull'usabilità dei siti web, parte dalle regole euristiche di Nielsen arrivando a conclusioni nuove ed, in alcuni casi, diverse. Secondo la sua teoria, grande importanza rivestono i sensi, e pure la memoria gioca un ruolo fondamentale nell'usabilità. Le conoscenze (informazioni)

fondamentali, affinché un utente sia in grado di interagire con un sistema e di usarlo, possono avere due collocazioni: il mondo (cioè l'interfaccia) e la mente (cioè la memoria). Si dovrà cercare di massimizzare la facilità con cui gli utenti ricorderanno le modalità di utilizzo del sito stesso e di minimizzare il loro carico mnemonico.

Pearrow suggerisce le seguenti norme:

- **Non creare pagine orfane**, cioè pagine che diventano vicoli ciechi.
- **Mantenere gli elementi costanti**: ogni pagina del sito deve avere delle comunanze che aiutino l'utente a identificare il sito in cui si trova.
- **Verificare il proprio lavoro in molti browser e condizioni**.
- **Pensare in modo globale**, così che il contenuto sia accessibile agli utenti che parlano diverse lingue.
- **Avvertire delle dimensioni di un file di cui si consente il download**.
- **Utilizzare la parte superiore e la parte sinistra dello schermo per l'esplorazione**.
- **Fornire un sistema di esplorazione supplementare**. Visto che le persone apprendono in modi diversi e hanno modi di esplorazione diversi. Si devono fornire le tre modalità di base per l'esplorazione: i collegamenti in ogni pagina, una mappa del sito e un motore di ricerca.
- **Eseguire test di usabilità sui siti della concorrenza**. Questa operazione può offrire il grande vantaggio di non commettere gli stessi errori.
- **Utilizzare URL facili da capire**, infatti creare nomi di directory che facciano riferimento ai concetti contenuti può aiutare le persone a orientarsi.
- **Non far perdere le tracce del sito**, se occorre collegarsi a pagine web che non fanno parte del sito.
- **Evitare colori altamente saturati**: i colori molto saturati sono positivi perché attirano l'attenzione per la loro luminosità, ma diventano a lungo andare stancanti.
- **Non esagerare troppo con i tipi di carattere**. Meno tipi di carattere ci sono, meglio è, perciò la cosa migliore è scegliere un carattere per le intestazioni e uno per il corpo del testo (entrambi sans serif).
- **Evitare i frame**

- **Evitare di aprire nuove finestre**, poiché l'apertura di nuove finestre del browser lascia gli utenti perplessi o li infastidisce.
- **Usare parsimonia**, poiché togliere o aggiungere elementi può sovvertire il significato del sito.

Il web distribuisce i suoi contenuti in maniera ipertestuale. L'ipertesto, in quanto medium, richiede una sua specifica modalità per la trasmissione del messaggio. Non si può quindi pensare ad un adattamento dello stile di scrittura su carta al web. Studi condotti sul comportamento dei navigatori hanno evidenziato che gli utenti di Internet scorrono le pagine web senza leggere ogni singola parola, che cercano termini o frasi particolari e solo se le individuano iniziano a leggere; che se essi non trovano velocemente e facilmente le informazioni specifiche cercate, passano oltre senza aspettare; e inoltre che l'uso di un gergo o di contenuti ambigui confonde poi gli utenti, rallentandoli e distraendoli. Infine si stima che la maggior parte delle persone legge il 25% più lentamente dalle pagine web che dalle pubblicazioni a stampa poiché la lettura a schermo è più difficile per gli occhi⁵.

In considerazione di quanto detto sopra, un ipertesto deve innanzitutto essere conciso, scorrevole e oggettivo e queste tre peculiarità possono essere conseguite per mezzo di alcuni accorgimenti. Nielsen⁶, a questo proposito, suggerisce:

- La sintassi deve essere costituita da paragrafi brevi, sottotitoli, liste puntate (che permettano la scorribilità del testo).
- Ogni paragrafo deve esprimere una sola idea riportando prima i concetti più importanti, secondo il principio della piramide capovolta.
- Quando è necessario fornire grandi quantità di informazioni è più efficace spezzettarle su più pagine creando una struttura gerarchica di rimandi dove vengano evidenziate ed enfatizzate le parole più importanti.
- I titoli devono essere pensati in maniera diversa rispetto agli altri mezzi di comunicazione. Mentre per la stampa e la televisione grande importanza

⁵ STEVE LEE, "Writing for the Web – How to Write Web-Friendly Content for your Site" <<http://www.freepint.com/issues/050603.htm#feature>>.

⁶ JACOB NIELSEN, *op. cit.*

hanno i titoli d'effetto, sul web essi devono rispondere ad una finalità esplicativa, devono cioè costituire un riassunto estremamente conciso del macrocontenuto, senza giochi di parole né doppi sensi. Il concetto o la persona di cui si parla nel resto della pagina devono essere espressi nel titolo.

Per quanto riguarda l'aspetto grafico del testo, Nielsen suggerisce di prestare attenzione alla dimensione dei caratteri e al tipo di font. È meglio scegliere una dimensione media per i caratteri, sfruttando i caratteri grandi per le parti che si vogliono sottolineare ed evitando caratteri troppo piccoli che possano mettere in difficoltà chi legge. È stato dimostrato che nei testi con una dimensione inferiore a 9 pt i caratteri senza grazie sono più leggibili e sforzano meno la vista, mentre nelle parti di testo di dimensioni maggiori i caratteri con le grazie danno una visione migliore.

Anche Steve Lee⁷ suggerisce di:

- ridurre approssimativamente della metà il numero delle parole usate per la stampa; evitando periodi introduttivi e forme passive;
- creare rimandi, per evitare inutili ripetizioni, se un'informazione è già presente altrove;
- utilizzare parole e frasi brevi e semplici (per esempio "usare" invece di "utilizzare");
- evidenziare i testi (facendo uso, per esempio, del grassetto o del corsivo, è meglio evitare le sottolineature per non creare confusione con i link) per portare l'attenzione del lettore sulle parole chiave;
- adottare un tono diretto e colloquiale nello scrivere tutti i testi.

Lee invita infine il creatore dell'ipertesto a chiedere a qualcuno di leggere con attenzione il testo della prima stesura o almeno di leggerlo lui personalmente ad alta voce per verificare che sia chiaro e ben costruito.

LA QUALITÀ DEI SITI CON FINALITÀ CULTURALI

Oltre ai principi di usabilità, nella costruzione di un sito web che abbia finalità culturali, devono essere fissati dei principi di qualità. Uno dei working group del

⁷ STEVE LEE, *op. cit.*

progetto europeo Minerva ha preso in considerazione questo aspetto, arrivando all'elaborazione di una serie di punti⁸:

- il sito deve **essere trasparente**, definendo chiaramente sia l'identità e gli obiettivi del sito web sia l'organismo responsabile della sua gestione;
- bisogna selezionare, digitalizzare, indicizzare, presentare e controllare i contenuti per creare un sito web **efficace** per tutti gli utenti;
- si devono **implementare linee guida per le politiche di qualità del servizio** per assicurare che il sito web venga adeguatamente mantenuto e aggiornato;
- il sito deve essere **accessibile** a tutti gli utenti, indipendentemente dalla tecnologia utilizzata o dalle loro inabilità, inclusi gli strumenti di navigazione, il contenuto e gli elementi interattivi;
- deve essere **centrato sull'utente**, tenendo conto delle sue esigenze, garantendo pertinenza della risposta e facilità d'uso attraverso meccanismi di valutazione e feedback;
- deve essere **reattivo**, consentendo agli utenti di contattare il sito e ricevere un'adeguata risposta. Se necessario, incoraggiare i quesiti, la condivisione dei dati e la discussione con e tra gli utenti;
- è necessario essere consapevoli dell'importanza del **multilinguismo** fornendo un livello minimo di accesso in più di una lingua;
- deve essere **interoperabile** all'interno delle reti culturali per consentire agli utenti di localizzare facilmente i contenuti e i servizi che rispondono alle loro necessità;
- deve essere **gestito** nel rispetto delle norme legali come il diritto di proprietà intellettuale e la riservatezza e indicare chiaramente i termini e le condizioni di utilizzo del sito web e dei suoi contenuti;
- è opportuno adottare strategie e standard per assicurare che il sito web e i suoi contenuti vengano **conservati** a lungo termine

⁸ MINERVA WORKING GROUP, Linee guida sulla costruzione dei siti culturali, novembre 2003
<<http://www.minervaeurope.org/structure/workinggroups/userneeds/documents/cwgp-i.htm>>.

LA STRUTTURAZIONE DEL SITO DEL PROGETTO BENUSSI

Come precedentemente spiegato, l'interfaccia è la parte necessaria per consentire all'utente di collegarsi al server remoto di ALEPH in cui sono conservate le descrizioni dei documenti. Attraverso la consultazione dei record archivistici l'utente avrà la possibilità di collegarsi ad un altro server per visualizzare i digital objects, cioè i fascicoli dell'archivio, contenenti le immagini in formato JPEG. L'interfaccia web del Progetto Benussi, consultabile all'indirizzo <<http://www.biblio.unimib.it/benussi.html>>, è stata costruita in linguaggio HTML; tale linguaggio non è però sufficiente a ricercare e presentare i risultati di un sistema di gestione delle immagini. È necessario creare una connessione delle pagine web al server remoto che gestisce il database dove si trovano i record di descrizione. Il record archivistico rimanda l'utente ad un altro server su cui si trovano i documenti digitali. I file sono identificati tramite un URN⁹, indicante il percorso per raggiungerli, che viene richiamato dal record.

⁹ L'URN, Uniform Resource Name, identifica una risorsa o un'unità di informazione indipendentemente dalla sua localizzazione. I requisiti funzionali sono stati creati dal gruppo di lavoro RCF 1737, che ne spiega l'utilità in questo modo:

" The resource identified by a URN may reside in one or more locations at any given time, may move or may not be available at all."

La sintassi di un URN indica esplicitamente il naming scheme includendo un identificatore. Un esempio di naming scheme è costituito dal Path, cioè dal percorso per raggiungere la risorsa, e da una stringa univoca, che indica il nome del file (/A/B/C/doc.html).

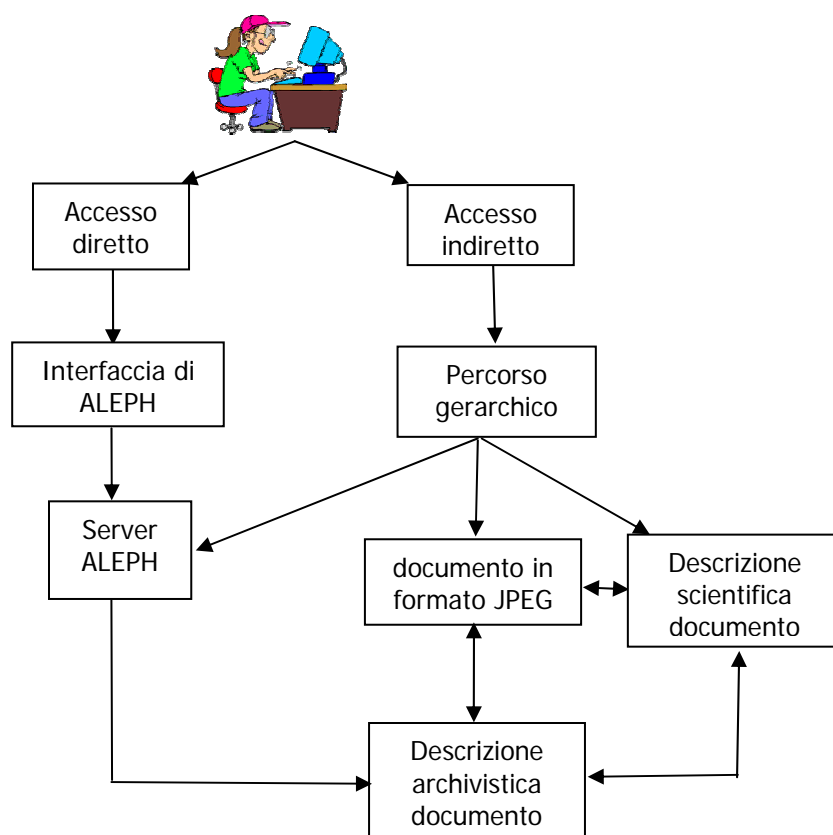


Figura 1: Rappresentazione della ricerca

La strutturazione e la grafica del sito sono state pensate e realizzate con la massima semplicità nell'intento di dare rilievo ai contenuti.

Nell'homepage sono stati inseriti pochi elementi: una barra di navigazione che rimanda a tutte le altre pagine del sito (Biografia, descrizione dell'archivio, ricerca, navigazione, tour attraverso i documenti più significativi, informazioni riguardanti il progetto, termini d'uso) ed il logo dell'Università degli Studi di Milano – Bicocca. Il colore dello sfondo è stato campionato dal documento, uno dei ritratti di Benussi disegnato da Gino Parin, che si vede in trasparenza. Le scritte di color rosso intenso richiamano il colore delle Università milanesi.



Figura 2: L'homepage

Nella parte alta di ogni pagina è presente una barra di collegamento a tutte le altre, mentre alla fine di ogni pagina sono riportati una barra di navigazione testuale, il simbolo dell'Università ed il simbolo del progetto di dimensioni inferiori rispetto a quelli dell'homepage. Il colore di fondo è il medesimo dell'homepage; per il corpo del testo è stato scelto il Verdana, un carattere senza grazie che sembrava rendere più semplice la lettura del testo, di dimensione 3, pari a 12 pt.

Il sito è stato ottimizzato per una risoluzione di 1024x678 e per il browser Internet Explorer, versione 5 e 6.

Nella pagina "BIOGRAFIA" è stato inserito uno dei ritratti di Benussi disegnato da Gino Parin. Vengono fornite informazioni sulla vita dello psicologo triestino e, per dare un'immagine più completa della sua personalità, si è pensato di inserire aneddoti, curiosità e descrizioni fatte da amici, allievi e colleghi.

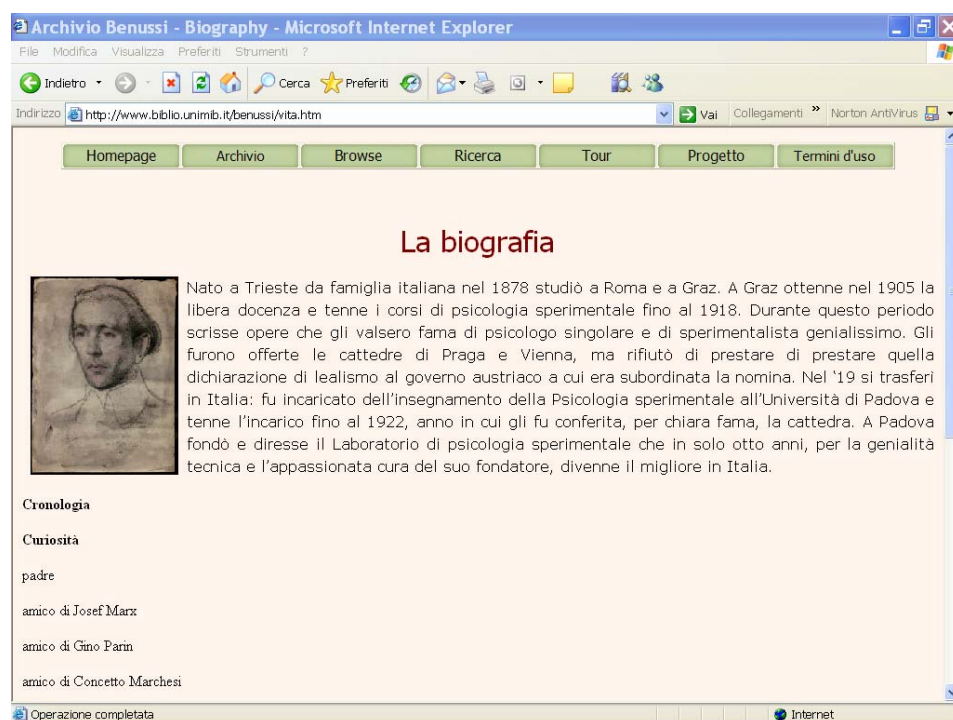


Figura 3: Pagina dedicata alla vita

Nella pagina "ARCHIVIO" viene presentata la descrizione a livello generale del fondo, in base alle regole ISAD(G).

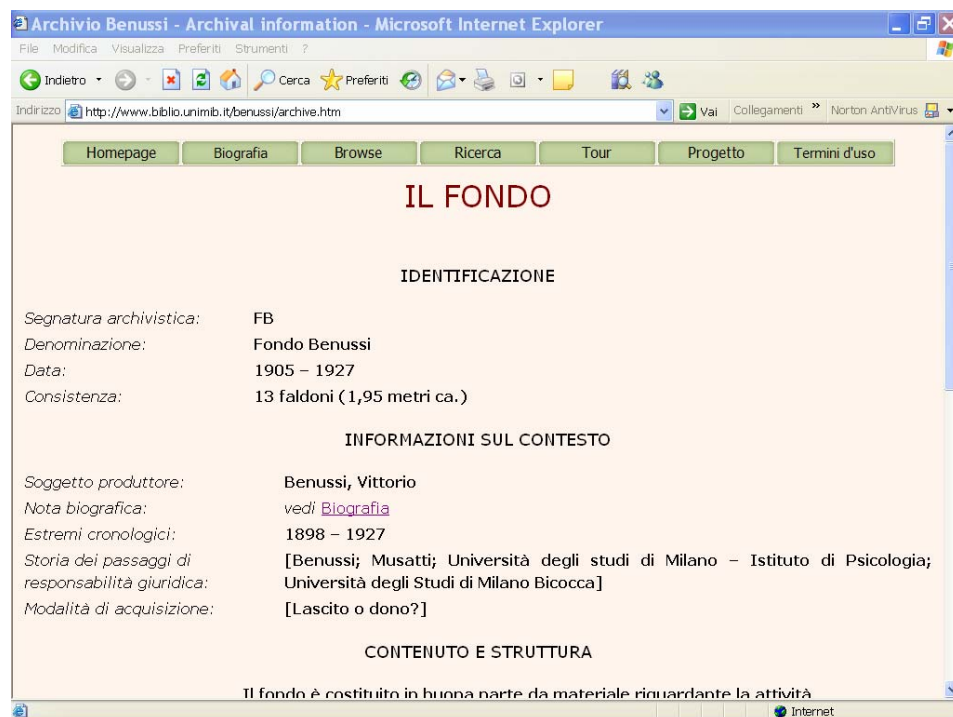


Figura 4: Pagina dedicata alla descrizione dell'archivio

Un discorso particolare e più approfondito meritano le pagine di ricerca e reperimento delle informazioni. La possibilità di ricercare e reperire facilmente la

documentazione, grazie ad un efficace sistema di metadati, differenzia un'aggregazione di immagini da un archivio digitale. L'accessibilità ai documenti dipende dalla quantità e dal tipo di metadati usati, che devono essere copiosi e ben strutturati in modo da offrire all'utente finale maggiori opzioni per il reperimento. L'efficacia del sistema di reperimento dipende anche e soprattutto dal modo in cui le informazioni vengono ricercate.

È sempre difficile anticipare i comportamenti degli utenti, ma in generale si può ipotizzare che ci saranno utenti che, sapendo già esattamente cosa vogliono, compiono ricerche specifiche; d'altra parte ci saranno utenti che, sapendo in linea generale quale argomento interessa loro, effettuano ricerche generiche; e ci saranno infine utenti che scorrono l'archivio in base alla filosofia della serendipity, cioè del trovare per caso ciò che interessa loro.

Le strategie di ricerca e la loro efficacia sono influenzate dalle conoscenze pregresse degli utenti; in genere l'utente inesperto si affida alla ricerca per parola, mentre quello che ha già una certa esperienza si affida a sistemi diversi. Alcune ricerche hanno individuato diversi tipi di comportamento nella ricerca di informazioni¹⁰:

- prescrittivo: usato per includere esigenze e restrizioni;
- esplorativo: tipicamente usato prima che una particolare direzione sia stata presa;
- finalizzato: è un metodo di ricerca più diretto ed informato;
- associativo: è una ricerca attiva di informazioni collegate e interconnesse;
- intuitivo: l'utente è diretto da sensazioni non specifiche;
- curioso: l'utente è alla ricerca di qualcosa che susciti interesse;
- tangenziale: l'utente va chiaramente oltre le esigenze iniziali;
- accidentale: l'utente compie azioni accidentali o deviazioni del sistema verso luoghi non premeditati.

I metodi di ricerca, oltre che dall'esperienza dell'utente, sono influenzati dalla finalità della ricerca, cioè dalle motivazioni che spingono l'utente a compierla.

¹⁰ TECHNICAL ADVISORY SERVICE FOR IMAGES "Searching for and Retrieving Digital Images" in *Technical Advisory Service for Images – Advice Paper*
<<http://www.tasi.ac.uk/advice/creating/pdf/srcandret.pdf>>.

Per questa diversità di approccio è necessario creare un'interfaccia integrata che si adatti alle esigenze di tutte le tipologie di utenza. Si possono impostare diversi tipi di ricerca:

- Ricerca libera. La ricerca viene compiuta in tutti i campi del record di descrizione, come titolo o parole chiave, ma ha due inconvenienti principali: la grossa quantità di risultati di diversa rilevanza (rumore) e la lentezza nella visualizzazione degli stessi.
- Ricerca per campi. Permette di ricercare più termini contemporaneamente, combinandoli attraverso l'uso degli operatori booleani AND, OR, NOT, e specificando in quali campi del record devono essere cercati ed in base a quale relazione logica.
- Ricerca per parole chiave. Le parole chiave sono parole che sono state selezionate per la loro particolare rilevanza o il loro particolare significato. I termini sono spesso presi da vocabolari controllati o da altre liste di intestazioni per soggetto. Questo tipo di ricerca offre risultati più specifici e di grande rilevanza rispetto alla parola cercata. Può avere due forme: può essere un sistema libero che non offre delle indicazioni oppure un sistema pilotato grazie alla presenza di sistemi di supporto come thesaurus, liste di intestazione, liste di soggetti in ordine alfabetico. Quest'ultimo tipo di organizzazione risulta molto vantaggioso per l'utente (soprattutto se non ha informazioni sull'archivio) ma richiede un lungo lavoro di elaborazione a monte.
- Ricerca attraverso un thesaurus. Fornisce una lista gerarchica di termini di ricerca alternativi ed è tipicamente basata su termini derivati da un vocabolario controllato. Questo tipo di ricerca può aumentare l'accuratezza e diminuire il rischio di reperire informazioni scarse o poco rilevanti. Normalmente capita che l'utente non ottenga risultati rilevanti perché non usa i termini giusti: aiutandosi con un thesaurus l'utente è sicuro di usare i termini adottati nella descrizione dei documenti.
- Ricerca predefinita. Esaminando le registrazioni dei termini usati dagli utenti è possibile identificare le parole o le frasi più ricercate e quindi fornire le ricerche già svolte come opzioni in una lista.

- Ricerca per scorrimento. Lo scorrimento è un metodo per trovare l'informazione; in generale si tratta di una struttura ad albero, gerarchicamente organizzata, di intestazioni per soggetto che va dal generale al particolare.
- Ricerca per raggruppamenti logici. Per un accesso facile e veloce, si può decidere di raggruppare i documenti in raccolte logiche. Questi raggruppamenti per esempio possono essere basati sul soggetto o sul periodo cronologico.

Nel progetto Benussi sono stati combinati quattro tipi di ricerca: la ricerca base, la ricerca per campi, la ricerca per liste e la ricerca per scorrimento (o browsing). Le prime tre opzioni sono offerte dall'interfaccia di ALEPH, mentre la ricerca per scorrimento è stata creata a parte riproducendo la struttura dell'archivio.

The screenshot shows the 'Fondo Benussi - Ricerca base' web interface. At the top, there's a navigation bar with links: 'Catalogo Benussi', 'Ricerca base', 'Ricerca per scorrimento', and 'Ricerca avanzata'. Below this is a search form with the following fields and options:

- Digita parola/e**: A text input field.
- Campo da ricercare**: A dropdown menu set to 'Tutti i campi'.
- Parole adiacenti?**: Radio buttons for 'No' (selected) and 'Si'.
- Limita la ricerca a:**
 - Lingua del documento**: A dropdown menu set to 'Tutte'.
 - Dall'anno**: A text input field.
 - All'anno**: A text input field.

Below the search form, there's a section titled 'Suggerimenti per la ricerca:' with several tips:

- Se scegli 'Parole adiacenti', puoi digitare la stringa *computer programming* e il sistema comprenderà che vuoi solo i record che contengono la parola *computer* ACCANTO alla parola *programming*.
- Una volta impostati dei limiti per la ricerca, questi vengono mantenuti anche nelle altre tipologie di ricerca e nelle ricerche successive.
- E' indifferente usare lettere maiuscole o minuscole. Ad esempio, *computer* recupererà sia *computer*, *Computer* che *COMPUTER*.
- L'operatore booleano AND viene assunto di default tra le parole. Puoi usare anche OR e NOT nella stringa di ricerca. Ad esempio, se digiti (*heart OR cardiac*) AND *surgery* il sistema recupererà tutti i record che contengono *heart* e *cardiac*, insieme alla parola *surgery*.
- Usa il carattere ? per mascherare parti di parola. Ad esempio, *gun?* recupererà *gun*, *guns*, *gunners*, *gunney*, *gunning*, etc. Un altro esempio: *?ology* recupererà *anthropology*, *archaeology*, *psychology*, ecc. Il carattere ? può essere usato per recuperare varianti ortografiche. Ad esempio, *alum?m* recupererà l'ortografia americana, *aluminum*, e quella britannica *aluminium*.

Figura 5: Interfaccia per la ricerca base

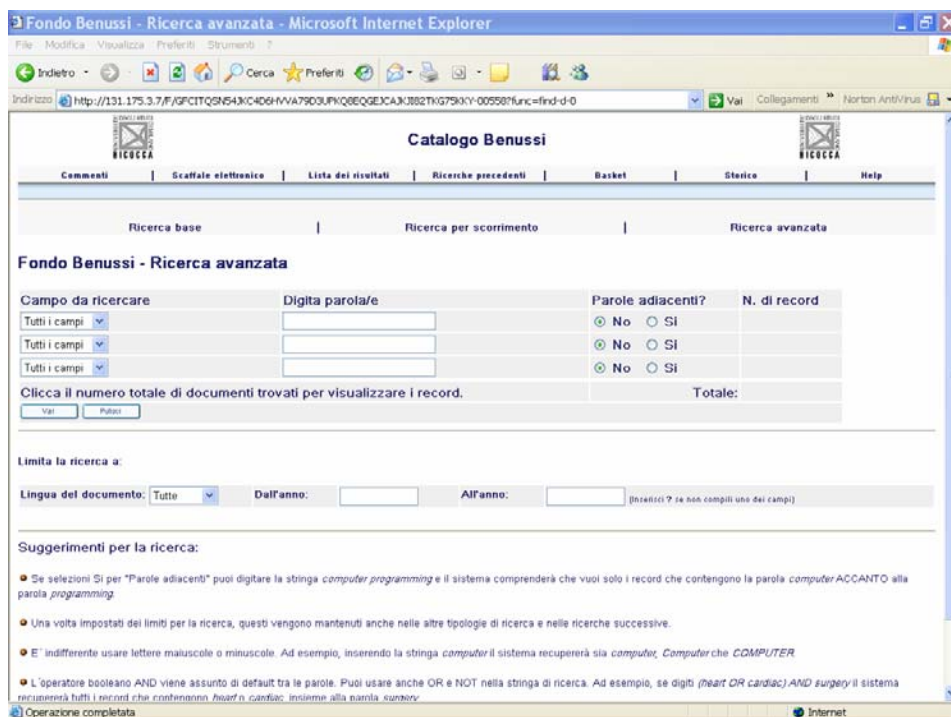


Figura 6: Interfaccia della ricerca per campi

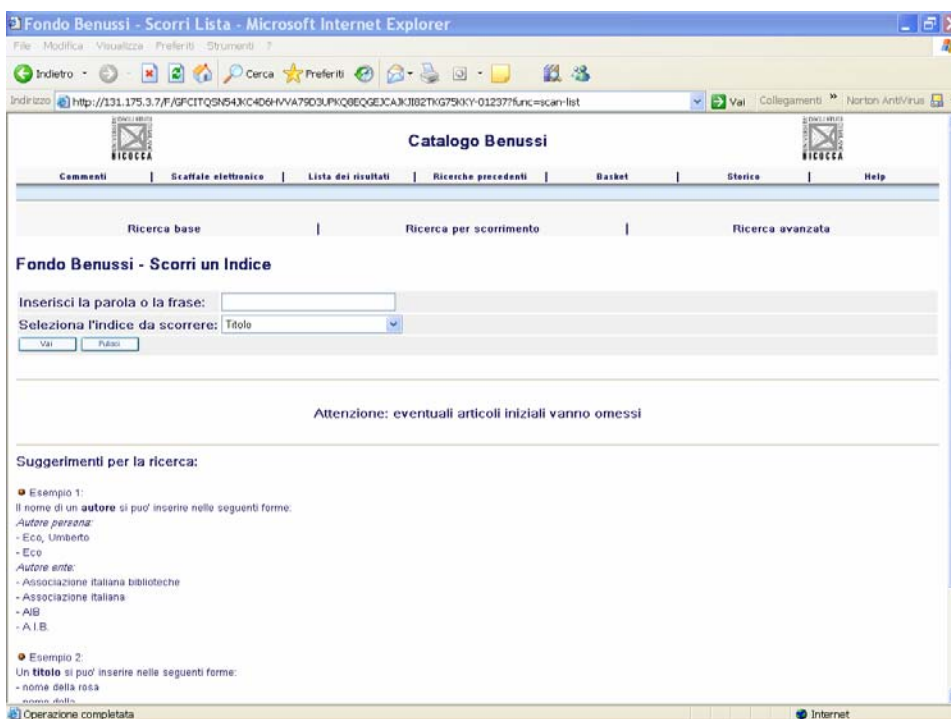


Figura 7: Interfaccia della ricerca per liste

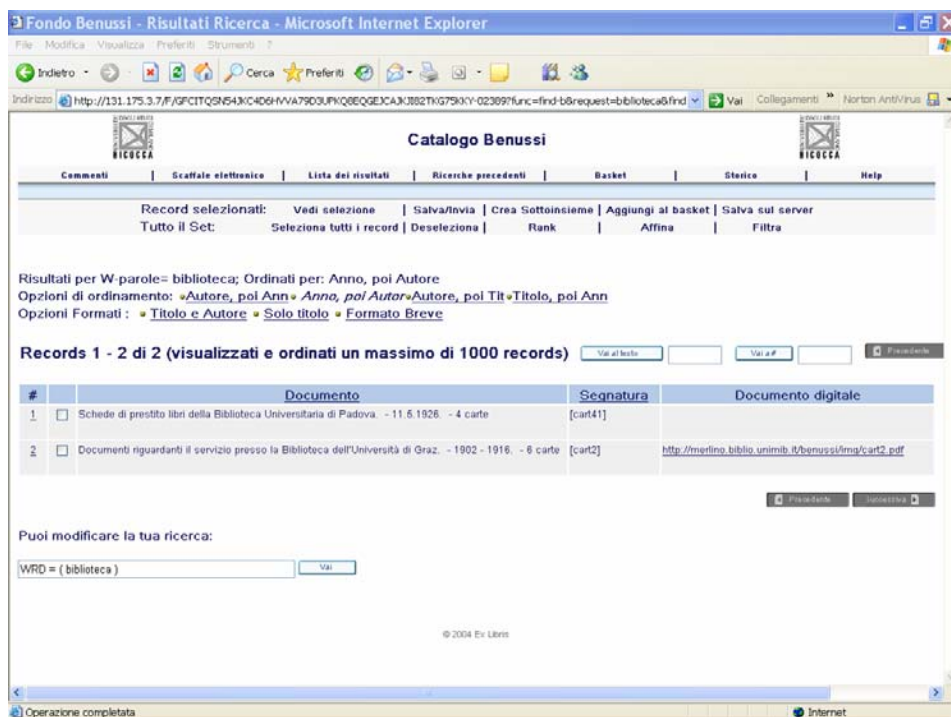


Figura 8: Short view dei risultati

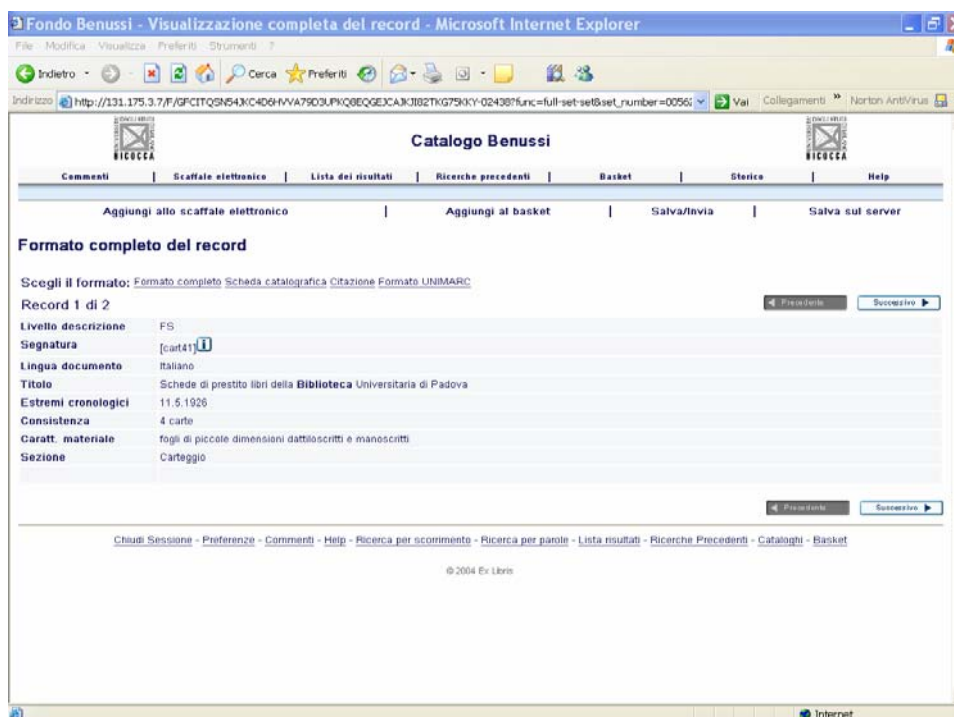


Figura 9: Long view di un risultato

L'interfaccia di ALEPH della versione 16.1 è stata personalizzata per meglio adattarsi alle esigenze del progetto. Dalla visualizzazione breve, dalla visualizzazione lunga e dalla navigazione (che si vede successivamente) è

possibile accedere al documento digitale e ad un piccolo riassunto che descrive e commenta il contenuto del documento stesso.

Il VIEWER per i documenti è stato realizzato attraverso l'uso di un javascript che permette di scorrere le pagine avanti ed indietro, di saltare da una pagina all'altra, di ingrandire le immagini fino ad un massimo di tre volte rispetto alla loro dimensione originale. I numeri al di sopra dell'immagine aiutano l'utente a sapere il punto del documento in cui si trova e di quante pagine è composto il documento. Questa è l'unica pagina in cui si è deciso di utilizzare un colore diverso per lo sfondo. La motivazione sta nel fatto che il colore usato per le altre pagine era troppo simile a quello dei documenti e la mancanza di una contrapposizione tra i colori non dava risalto alle immagini rendendo più complicata la lettura dei documenti. Per non stravolgere totalmente lo stile del sito si era pensato, in un primo momento, di creare un negativo delle altre pagine invertendo il colore dello sfondo con quello delle intestazioni, cioè di realizzare lo sfondo in rosso e le intestazioni in beige; il rosso intenso scelto per le intestazioni risultava però fastidioso se usato come sfondo per l'intera pagina. Si è optato perciò per un colore sempre scuro, ma più rilassante per la vista e che si adattasse al colore dei bottoni e del testo.

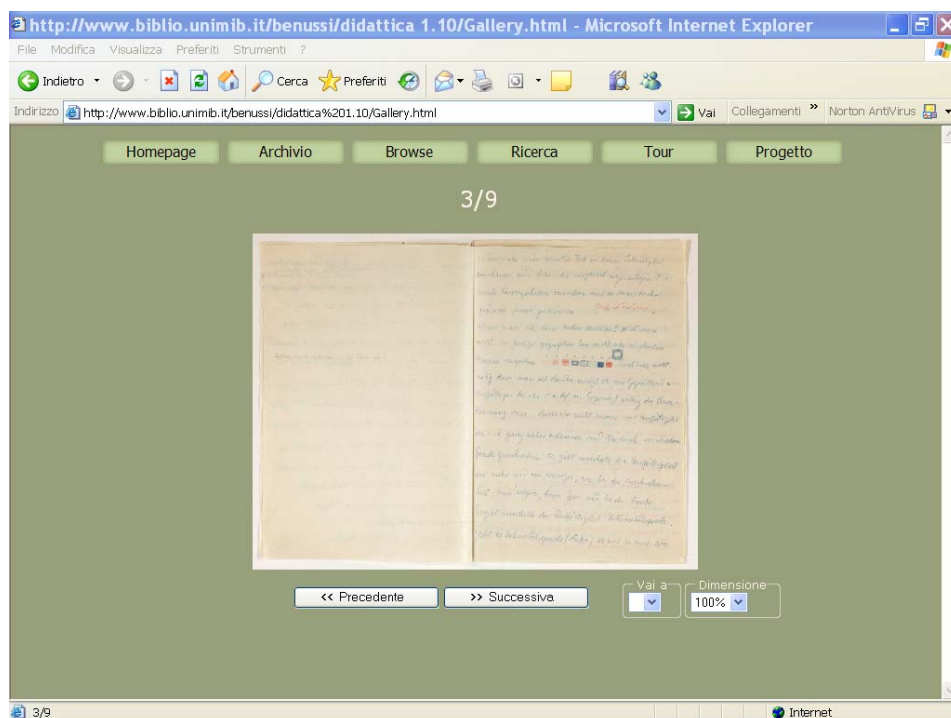


Figura 10: Il viewer per i documenti

Per il commento dei documenti sono state create altre pagine in formato HTML collegate al record di descrizione in ALEPH, alla pagina in cui è rappresentato il percorso gerarchico e alla pagina viewer.

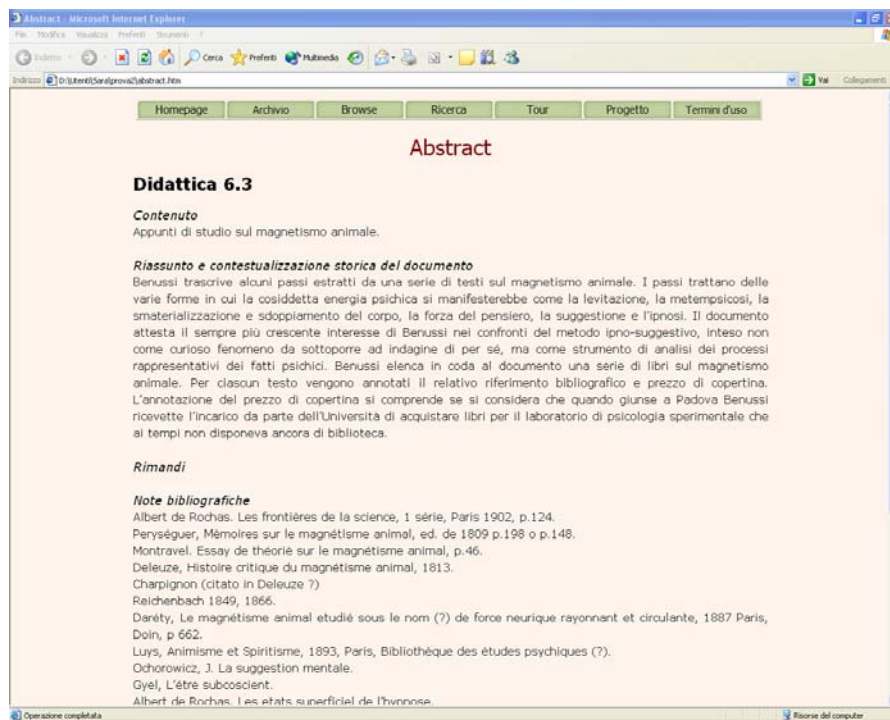


Figura 11: Ogni documento è già stato o sarà fornito di un suo commento

La costruzione della pagina relativa alla “NAVIGAZIONE” ha creato maggiori difficoltà a livello progettuale.

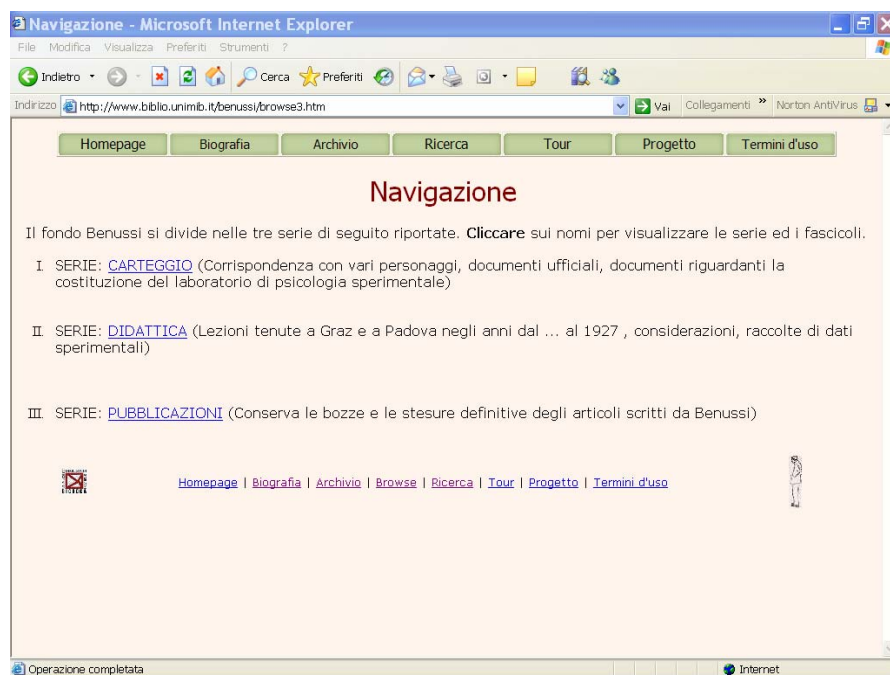
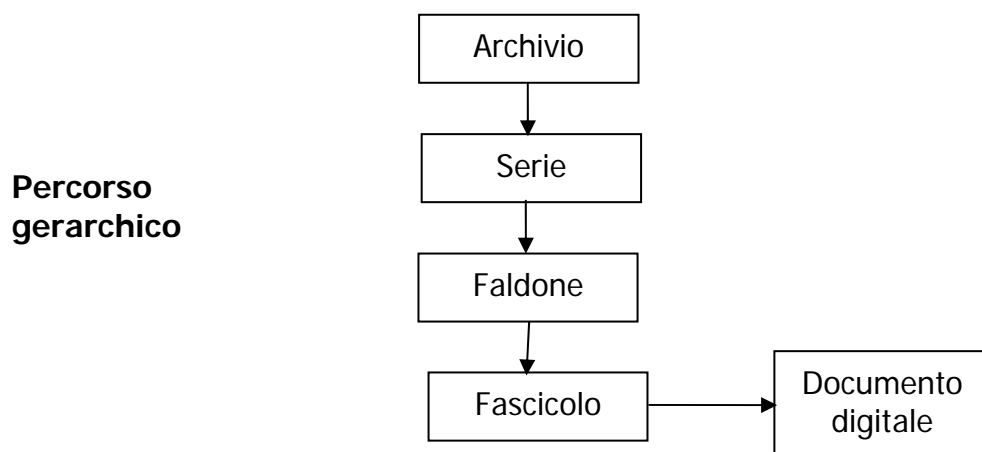


Figura 12: Interfaccia per la navigazione dell'archivio

L'uso del web ha cambiato il modo di rapportarsi con gli archivi e le possibilità di ricerca al loro interno. Oltre agli specialisti, agli storici, oggi capita anche ai surfisti del web di entrare negli archivi digitali e di usarne il materiale. Questo comporta la necessità di trovare nuovi strumenti di mediazione e corredo per la struttura dell'archivio. Come afferma Stefano Vitali:

I nuovi mezzi di comunicazione non vanno d'accordo con i vecchi metodi, o ci vanno solo parzialmente, poiché le condizioni stesse della comunicazione sono difformi. [...] La pubblicazione in linea di strumenti di ricerca non può fare affidamento sulla mediazione fra documentazione, strumenti ed utente che è assicurata dall'interazione 'faccia a faccia' fra archivista e ricercatore. [...] l'informatica e ancor più la comunicazione via rete costringe a rendere esplicito ciò che è implicito, a formalizzare flussi di conoscenze informali, a razionalizzare procedure 'spontanee'.¹¹

La finalità della struttura di navigazione dovrebbe essere quella di permettere all'utente di scorrere il contenuto dell'archivio attraverso un percorso gerarchico che consenta un approccio più tradizionale all'archivio. Il problema è stato come ricreare questo approccio di tipo gerarchico.



La difficoltà nella creazione di questa pagina nasce dallo sforzo di riuscire a creare uno strumento efficace e facilmente usabile da parte degli utenti senza

¹¹ STEFANO VITALI, "Archivi on-line: qualche riflessione metodologica", <<http://www.storia.unina.it/perfez/vitali.rtf>>.

l'intermediazione dell'archivista. La soluzione migliore è sembrata quella di creare, a livello di serie e di faldoni, degli elenchi a cascata in cui vengono visualizzati tutti i faldoni contenuti nelle serie

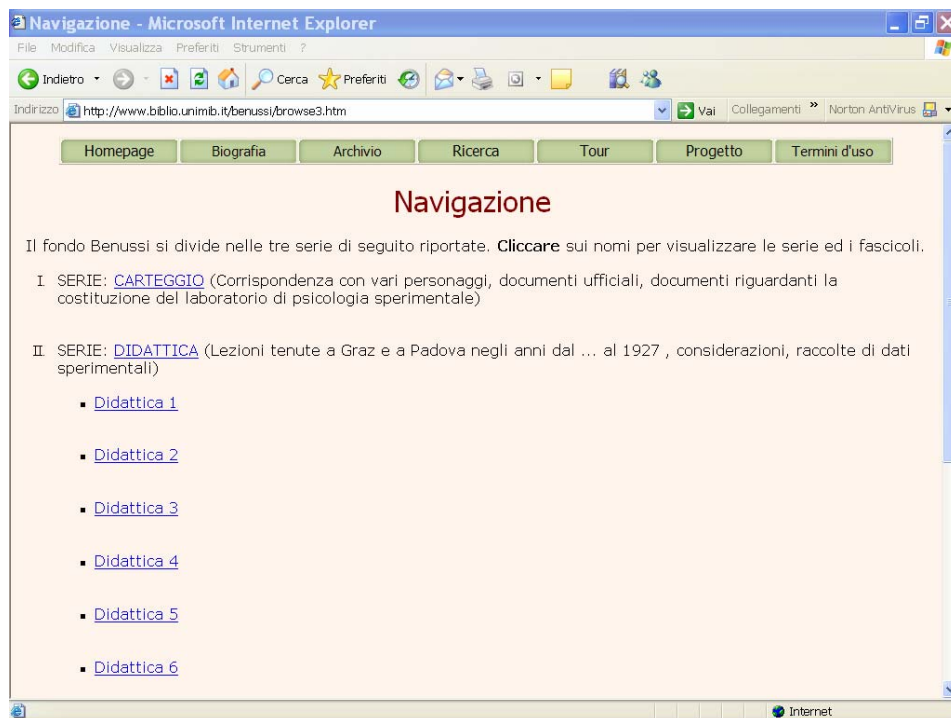


Figura 13: Interfaccia per la navigazione nell'archivio

e poi tutti i fascicoli contenuti nei vari faldoni.

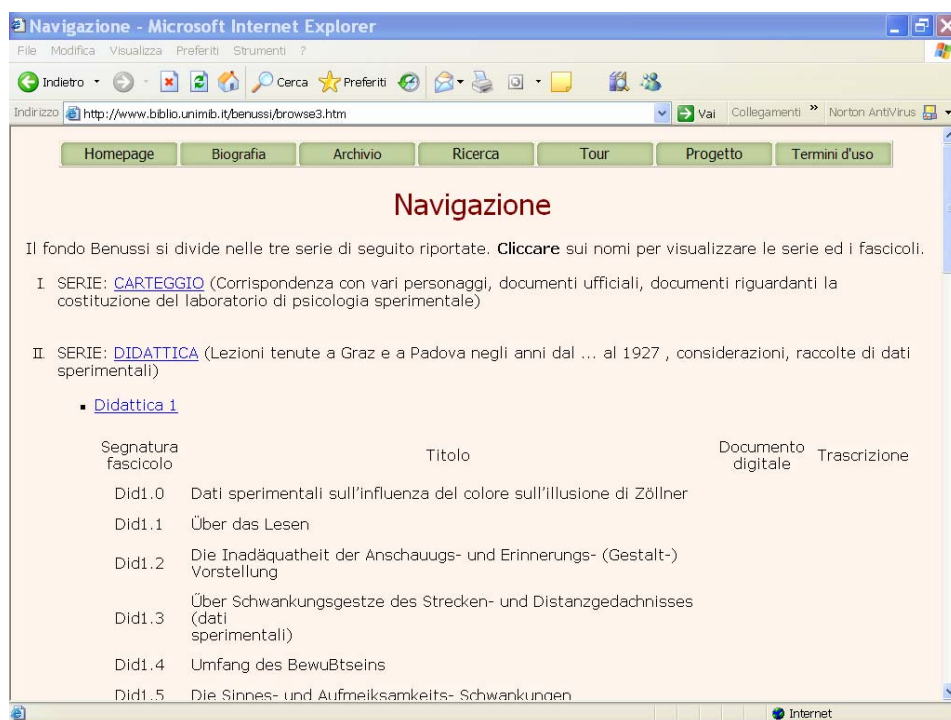


Figura 14: Interfaccia per la navigazione nell'archivio

Nelle sue riflessioni metodologiche Vitali continua:

Se si riflette sulle pratiche di lavoro negli Archivi – pratiche di ordinamento e descrizione – prevalse negli ultimi 100/150 anni, non si può non osservare come i principi fondamentali che le hanno guidate hanno costituito una sorta di 'precondizione metodologica implicita all'utilizzo della documentazione archivistica. [...] tali principi hanno teso a salvaguardare il contesto di origine e le modalità attraverso le quali la documentazione è pervenuta. Proprio perché il contesto di origine, e quindi il significato e il 'valore' originari, della documentazione è preservato e si presenta in forma auto-esplicativa è possibile per lo storico utilizzare criticamente le informazioni tratte dai documenti, ricomponendole in altri contesti. Le pratiche di ordinamento, nello sforzo di salvaguardare le strutture archivistiche originarie, ed i metodi descrittivi che tali strutture propongono di rendere trasparenti, rappresentano una critica implicita della documentazione, una precondizione. [...] L'ingresso dell'informatica e delle reti in questo settore ha complicato ulteriormente le cose, poiché non ha solo accentuato il ruolo della tecnologia quale mediatore fra produttore e ricettore delle informazioni, ma ha anche introdotto nuove metodologie descrittive, di cui si tratta di verificare più approfonditamente quanto conservino o quanto alterino quelle precondizioni implicite.¹²

In generale il successo di ogni ricerca per parola dipende dalla capacità di interrogare la banca dati o il motore di ricerca e dalla scelta dei vocaboli giusti. Un qualsiasi navigatore del web ha potuto molto spesso sperimentare cosa significa inserire termini poco indicati nella ricerca con Google: vengono restituiti numerosi riferimenti di cui, magari, nessuno veramente pertinente.

In ambito archivistico alle ambiguità usuali del linguaggio corrente si sommano quelle relative alla dimensione storicamente e geograficamente stratificata che dalla documentazione archivistica descritta si riflette all'interno delle relative descrizioni. Inoltre la ricerca per parola consente un recupero puntiforme dei singoli record, cioè offre in risposta singoli documenti estrapolati dal loro contesto. In questo modo si perde la percezione dell'archivio nel suo insieme e quindi il suo contesto di origine, molto importante come sottolineato in precedenza. Il pregio della navigazione dice ancora Vitali è quello di rendere più trasparente la struttura dell'archivio e quindi più consapevoli e meno misteriosi i percorsi di ricerca al suo interno.¹³

La pagina "TOUR" è stata ideata prendendo spunto dai musei virtuali e dalle esposizioni, di materiale particolare, organizzate negli spazi di Biblioteche ed Archivi. Si è ritenuto infatti che l'archivio contenesse materiale, più o meno inerente alla psicologia, interessante e ricco di curiosità, che probabilmente sarebbe sfuggito all'utente.

¹² STEFANO VITALI, *op. cit.*

¹³ STEFANO VITALI, *op. cit.*



Figura 15: Pagina d'introduzione ai tour

Sono stati organizzati tre percorsi guidati in cui sono presentati grafici, disegni, quadri, fotografie corredati da didascalie e rimandi al fascicolo in cui si trovano. Nel primo percorso sono stati inseriti i molti disegni di Benussi, tra cui un autoritratto, e diversi ritratti del pittore Gino Parin.



Figura 16: Tour 1

Nel secondo tour sono state inserite tutte le foto ed alcune cartoline.

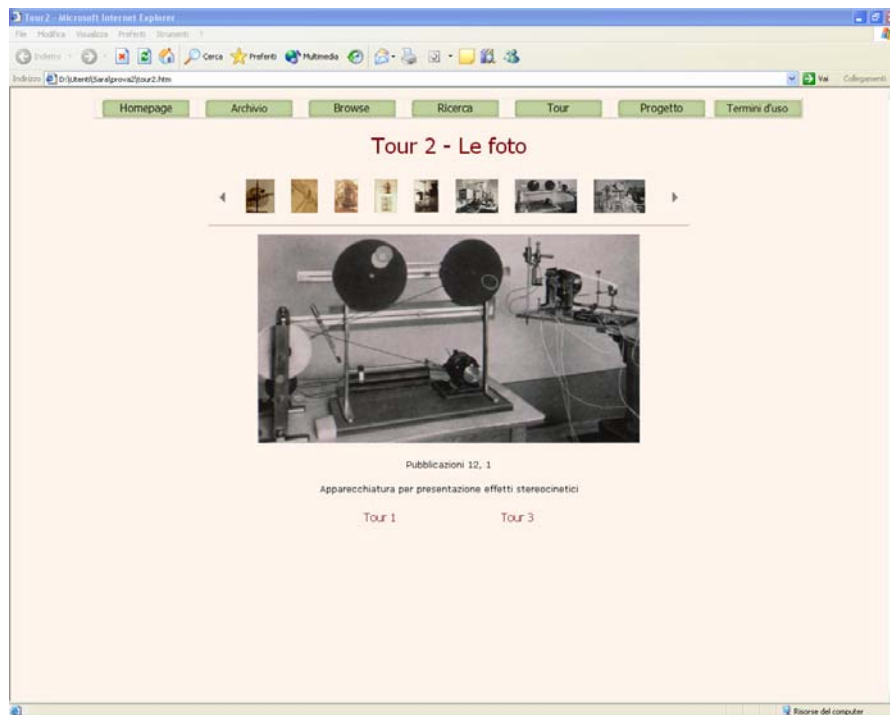


Figura 17: Tour 2

Il terzo percorso illustra l'archivio attraverso tutti i disegni, per lo più di progettazione, delle macchine sperimentali usate all'interno del Laboratorio.



Figura 18: Tour 3

Le "INFORMAZIONI SUL PROGETTO" sono state divise in tre gruppi, un primo gruppo con informazioni riguardanti i partecipanti al progetto, un secondo riguardante le informazioni e le specifiche tecniche ed una terza sezione dedicata agli eventuali sviluppi futuri. Vengono inoltre riportati i dati della Biblioteca ed il link al sito, i dati sul Dipartimento di Psicologia ed il collegamento alla sua pagina web e gli indirizzi e-mail di tutti i partecipanti al progetto per dare la possibilità di avere ulteriori informazioni.

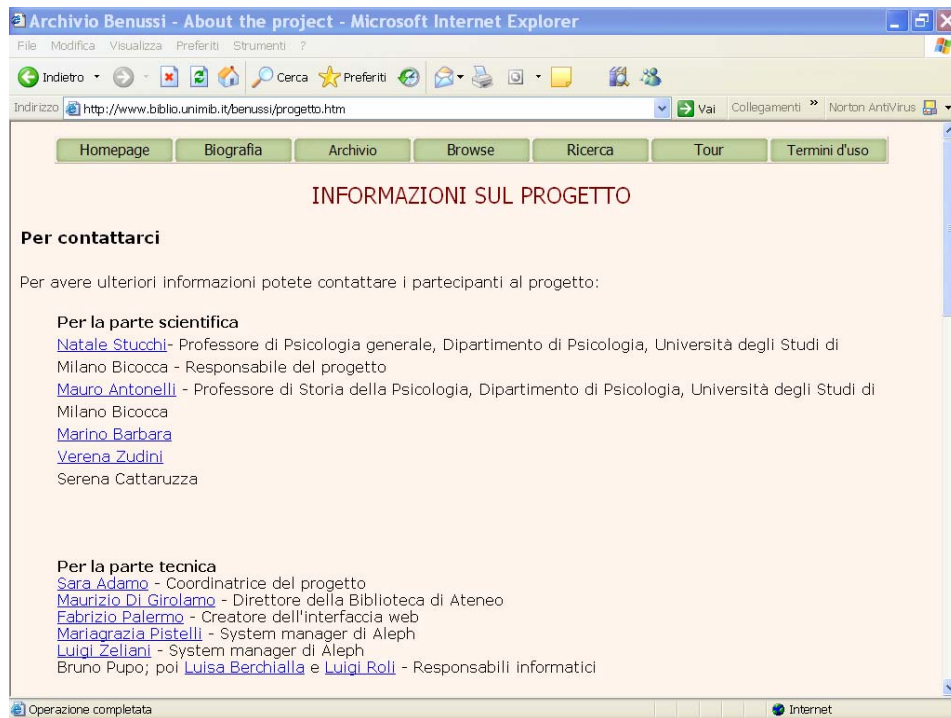


Figura 19: La pagina riporta le specifiche tecniche del progetto ed i contatti

I "TERMINI D'USO" presentano un'informativa legale sull'uso delle immagini digitalizzate e degli studi su tali immagini ed un form da compilare ed inviare per poterne fare un diverso uso.



Figura 20: Pagina sulle condizioni d'uso del materiale

L'ACCESSO

Secondo Giovanni Pascuzzi e Roberto Caso, per quanto riguarda la problematica dell'accesso in era digitale, tre fenomeni assumono grande rilevanza¹⁴: la possibilità di trasformare vari fenomeni analogici in altrettanti file di codice binario, quella di elaborare quei file e quella di trasmetterli velocemente e senza spese onerose in tutto il mondo attraverso Internet.

La sfida che l'era digitale porta ai modelli tradizionali di tutela del diritto d'autore presenta tre aspetti:

- l'estrema facilità di riproduzione delle opere;
- l'impossibilità di distinguere la copia dall'originale sul piano qualitativo;
- la facilità di distribuzione delle opere¹⁵.

Tali problematiche sono state ad esempio già largamente verificate dall'industria della musica.

¹⁴ GIOVANNI PASCUZZI, ROBERTO CASO, *I diritti sulle opere digitali: copyright statunitense e diritto d'autore italiano*, Padova, CEDAM, 2002, p. 10.

¹⁵ GIOVANNI PASCUZZI, ROBERTO CASO, *op. cit.*, p. 11.

Le vecchie leggi sul diritto d'autore non sono applicabili alle tipologie di beni concepiti con le nuove tecnologie informatiche, poiché queste cambiano il concetto di opera (non hanno più necessariamente un supporto materiale, possono mutare ed accrescersi in tempo reale, possono essere multimediali, cioè assemblare tecniche diverse), il concetto di autore (sempre più spesso non è il singolo a realizzare un progetto, che è il risultato della collaborazione di un team) ed il concetto di creatività (la digitalizzazione consente di creare nuovi contenuti attraverso la diversa combinazione di elementi e dati nuovi).

Nello sviluppo di un progetto di Biblioteca Digitale e di digitalizzazione di materiale un punto particolarmente critico è quindi costituito dalle norme di accesso e utilizzazione della stessa.

Non dimenticando che il fine di ogni biblioteca è quello di diffusione della cultura e di supporto alla ricerca scientifica, è giusto tutelare il materiale da essa posseduto da possibili manomissioni, snaturamenti ed usi del materiale a fini di lucro.

Per tutelare le immagini digitalizzate si può ricorrere sostanzialmente a tre strumenti: la legge sul diritto d'autore, l'elaborazione di un disclaimer che stabilisca le condizioni d'uso ed il ricorso a misure tecnologiche di protezione.

IL DIRITTO D'AUTORE

La tutela del diritto d'autore nel caso di un archivio digitale, come quello in esame, pone diverse questioni:

- a) La prima riguarda l'accesso, cioè la possibilità o meno di consultare i documenti presenti all'interno dell'archivio stesso.

Il decreto legislativo del 22 gennaio 2004, n. 41 all'articolo 122 recita:

I documenti conservati negli Archivi di Stato e negli archivi storici delle regioni, degli altri enti pubblici territoriali, nonché di ogni altro ente ed istituto pubblico sono liberamente consultabili, ad eccezione: di quelli dichiarati di carattere riservato ai sensi dell'art. 125, relativi alla politica estera o interna dello stato, che diventano consultabili dopo 50 anni dalla loro data; di quelli contenenti i dati sensibili, [...], che diventano consultabili 40 anni dopo la loro

data. Il termine è di settant'anni se i dati sono idonei a rivelare lo stato di salute, la vita sessuale o rapporti riservati di tipo familiare.¹⁶

Vittorio Benussi è morto nel 1927, dunque tutti i documenti all'interno del suo archivio risalgono almeno a quell'anno e sono tutti consultabili.

- b) L'archivio è composto in larga parte da materiale scientifico mai pubblicato. La disciplina sul diritto d'autore riconosce al creatore di un'opera prerogative economiche (diritto di copia) e morali (diritto morale). Mentre il diritto di copia decade dopo 70 anni, il diritto morale, cioè il diritto di rivendicare la paternità dell'opera opponendosi anche a mutilazioni e deformazioni della stessa, rimane sempre valido.
- c) Oltre che tutelare il contenuto dell'archivio in quanto creazione originale di Benussi è necessario tutelare il lavoro di digitalizzazione e di strutturazione dell'archivio in formato digitale.

L'esigenza di ampliamento della nozione di creatività si è presentata nel caso delle banche dati. Anche se un database non presenta una struttura originale, gli sforzi per la ricerca e la creazione dei contenuti spiegano la necessità di tutelare anche questo tipo di opere.

La principale convenzione internazionale sul Diritto d'autore rimane la Convenzione di Unione di Berna.

Di particolare rilievo riguardo ai problemi di Internet ed alle opere di ingegno è il WIPO Copyright Treaty (WCT)¹⁷. Tale trattato, che si appoggia in larga parte sul modello statunitense, si discosta dalla concezione europea del diritto d'autore. A livello di previsioni orizzontali si richiama la distinzione fondante tra idea (non protetta) e forma espressiva e si disciplinano i vari diritti esclusivi, le relative limitazioni e le eccezioni. Viene poi contemplata l'innovativa tutela delle misure tecnologiche e delle informazioni sul regime dei diritti. In merito alle banche dati il Trattato accorda la tutela alle compilazioni di dati o di altro materiale che costituiscano creazioni intellettuali per la selezione o la

¹⁶ Supplemento ordinario alla "Gazzetta Ufficiale della Repubblica Italiana", n. 45 del 24 febbraio 2004, Serie generale.

¹⁷ Il WIPO Copyright Treaty è stato sottoscritto a Ginevra nel 1996 dalla World Intellectual Property Organization. Il testo del trattato è scaricabile dal sito <http://ecommerce.wipo.int/survey/index.html>.

disposizione del loro contenuto. Nessuna tutela specifica è stata sinora prevista riguardo alla tutela del contenuto in sé. Novità introdotta dal Trattato è, in considerazione delle caratteristiche di Internet, il riconoscimento agli autori del diritto esclusivo di autorizzare ogni comunicazione al pubblico delle proprie opere sul filo o via etere; si prevede inoltre l'obbligo per le Parti contraenti di provvedere ad un'adeguata tutela giuridica, oltre che di preconstituire efficaci mezzi di soccorso, con riguardo alle misure tecnologiche usate dall'autore per proteggere i propri diritti e impedire atti non autorizzati.

A livello comunitario la direttiva 96/9/CEE mira ad offrire una doppia protezione: la copertura del diritto d'autore per quelle raccolte di dati che presentano una struttura originale ed un diritto sui generis, cioè un diritto di privativa al creatore sull'estrazione ed il riutilizzo dei dati¹⁸.

Nella legislazione della comunità europea si possono individuare tre filosofie principali¹⁹: la concezione economica del diritto d'autore, la ricerca di un equilibrio tra le prescrizioni del Trattato e la preoccupazione degli autori, la promozione di regole comunitarie innovative.

Uno dei risultati è il ravvicinamento o la convergenza tra disciplina delle opere dell'ingegno e disciplina delle invenzioni.

La direttiva CEE del 11 marzo 1996²⁰ definisce la banca dati come una raccolta di opere, dati o altri elementi indipendenti, sistematicamente o organicamente disposti, ed individualmente accessibili grazie a mezzi elettronici o in altro modo.

La direttiva 91/250/CEE²¹ riconosce il classico diritto d'autore unicamente per quanto concerne la struttura esterna della raccolta e non il suo contenuto. Perché una banca dati possa essere protetta dal diritto d'autore occorre che essa costituisca una creazione intellettuale propria dell'autore e che abbia una sua originalità nella scelta o nella disposizione del materiale.

La presunta insufficienza della sola protezione del diritto d'autore è stata controbilanciata dall'introduzione di un diritto speciale deputato essenzialmente

¹⁸ GIOVANNI PASCUZZI, ROBERTO CASO, *op. cit.*

¹⁹ *Ibidem*

²⁰ In Gazzetta Ufficiale della Comunità Europea, L 77, 27 marzo 1996, 20.

²¹ GIOVANNI PASCUZZI, ROBERTO CASO, *op. cit.*, p. 243.

a tutelare gli investimenti nella costituzione della banca dati, del tutto autonomo dalla tutela del diritto d'autore.

Il creatore gode di un ampio diritto di riproduzione, cioè del diritto esclusivo di effettuare o autorizzare la riproduzione permanente o temporanea, totale o parziale, con qualsiasi mezzo e in qualsivoglia forma; gode inoltre del diritto esclusivo di eseguire o autorizzare la traduzione, l'adattamento, una diversa disposizione e ogni altra modifica, nonché del diritto di eseguire o autorizzare qualsiasi distribuzione al pubblico dell'originale o di copie.

L'utente legittimo ha il diritto di riproduzione, traduzione, modifica etc. che sono necessari per l'accesso al contenuto della banca dati e per il normale impiego di essa. Viene dato il diritto di uso della banca dati per finalità di illustrazione didattica o di ricerca scientifica, nei limiti di quanto giustificato dallo scopo non commerciale perseguito.²²

Il diritto sui generis ha sostanzialmente il fine della tutela economica del produttore, ma il suo oggetto è la totalità o una parte sostanziale del contenuto della banca stessa (art. 7 par. 1); la fattispecie costitutiva di questo diritto è rappresentata da un investimento rilevante (finanziario, in termini di tempo, di lavoro e di energia), come quella del diritto d'autore è costituita dal requisito di originalità dell'opera.

Il costituente della banca dati ha in particolare il diritto di vietare operazioni di estrazione e/o reimpiego della totalità o di una parte sostanziale del contenuto della stessa, valutata in termini qualitativi (alla luce degli interessi dell'utilizzatore finale) o quantitativi.

Con il divieto di estrazione si intende vietare qualsiasi trasferimento permanente o temporaneo del contenuto su un altro supporto con qualsiasi mezzo o in qualsiasi forma. La seconda facoltà consiste nel vietare qualsiasi forma di messa a disposizione del pubblico della totalità o di una parte sostanziale del contenuto mediante distribuzione di copie, noleggio, trasmissione in linea o in altre forme.

La più recente direttiva comunitaria n. 2001/29/CEE concernente l'armonizzazione di taluni aspetti del diritto d'autore e dei diritti connessi nella società dell'informazione ha scelto di disciplinare in maniera orizzontale i diritti di riproduzione, comunicazione al pubblico e messa a disposizione del pubblico,

²² GIOVANNI PASCUZZI, ROBERTO CASO, *op. cit.*, p. 278.

distribuzione, e le relative eccezioni, nonché di apprestare un'efficace tutela delle misure tecnologiche e delle informazioni sul regime dei diritti.

La normativa estende a tutti gli autori di opere di ingegno il diritto esclusivo di autorizzare o vietare la riproduzione diretta o indiretta, temporanea o permanente, in qualunque modo o forma, temporanea o permanente. L'unica limitazione è data dalla libertà dell'utente di eseguire tutti gli atti di riproduzione privi di rilievo economico proprio, transitori o accessori, e parte integrante e essenziale di un procedimento tecnologico, eseguiti allo scopo di consentire la trasmissione in rete tra terzi con l'intervento di un intermediario o l'utilizzo legittimo di un'opera o di altri materiali.

La legislazione italiana si è uniformata alle disposizioni di legge della Comunità Europea. La legge che si occupa della protezione del diritto d'autore è la numero 633 del 22 aprile 1941, l'ultimo aggiornamento è stato effettuato con la legge 22 maggio 2004 n. 128.

L'articolo 1 della suddetta legge recita:

Sono protette ai sensi di questa legge le opere dell'ingegno di carattere creativo che appartengono alla letteratura, alla musica, alle arti figurative, all'architettura, al teatro ed alla cinematografia, qualunque ne sia il modo o la forma di espressione. Sono altresì protetti i programmi per elaboratore come opere letterarie ai sensi della convenzione di Berna sulle opere letterarie ed artistiche ratificata e resa esecutiva con legge 20 giugno 1978, n. 399, nonché le banche dati che per la scelta o la disposizione del materiale costituiscono una creazione intellettuale dell'autore.²³

All'articolo 2, comma 9 viene quindi spiegato ciò che la legislazione italiana intende per Banche Dati:

[...] raccolte di opere, dati o altri elementi indipendenti sistematicamente o metodicamente disposti ed individualmente accessibili mediante mezzi elettronici o in altro modo.²⁴

La tutela delle banche dati non fa quindi riferimento al contenuto, che rimane coperto dal suo diritto specifico, ma alla raccolta nell'insieme.

All'articolo 64- quinquies si legge che l'autore ha il diritto esclusivo di eseguire o autorizzare:

²³ Testo della Legge 22 aprile 1941, n. 633.

²⁴ *Ibidem*

- la riproduzione temporanea, totale o parziale, con qualsiasi mezzo ed in qualsiasi forma;
- la traduzione, l'adattamento, una diversa disposizione e ogni altra modifica;
- qualsiasi forma di distribuzione al pubblico dell'originale o di copie della Banca Dati;
- qualsiasi presentazione, dimostrazione o comunicazione in pubblico.

Il costituente della Banca Dati ha inoltre il diritto esclusivo di vietare le operazioni di estrazione ed ogni reimpiego della totalità o di una parte sostanziale della stessa. Tale diritto però sorge al momento del completamento della Banca Dati e si estingue trascorsi quindici anni dal 1° gennaio dell'anno successivo alla data della prima messa a disposizione del pubblico. Tale limite slitta di altri quindici anni solo in caso di cambiamenti sostanziali o integrazioni al contenuto tali da richiedere nuovi investimenti rilevanti (art. 102-bis).

DISCLAIMER

Il disclaimer è un atto di rinuncia ad un diritto. Solitamente consiste in una formula con cui l'ente creatore della biblioteca digitale esplicita all'utente in che modo può usare i documenti che gli vengono messi a disposizione. Generalmente l'utente ha la possibilità di salvare e stampare le immagini dei documenti digitalizzati per scopi privati e di ricerca, mentre l'uso pubblico del materiale o la pubblicazione su qualsiasi supporto avviene previa approvazione dell'ente creatore della biblioteca digitale. Il progetto Benussi si muove in questa direzione.

LE MISURE TECNOLOGICHE DI PROTEZIONE

Molto importante nella tutela del diritto d'autore risulta il ricorso a misure tecnologiche, come prospetta la direttiva comunitaria 2001/29/CE, "per impedire o limitare atti non autorizzati dal titolare del diritto d'autore"²⁵.

²⁵ Direttiva comunitaria 2001/29/CE, art. 47.

Largamente in uso sono i sistemi di marchiatura, basati su algoritmi matematici che forniscono una sorta di contrassegno, e gli identifier. Il loro utilizzo fa sì che un'opera posta in rete, precedentemente "marchiata" con codici, possa essere ugualmente riconoscibile e identificabile anche dopo interventi modificativi posti in atto da terzi²⁶.

La marcatura digitale ha fondamentalmente due funzioni, una è quella propria anche del timbro tradizionale sui libri, indicare il possesso del documento, l'altra è quella di proteggere il documento da eventuali manipolazioni. La marcatura avviene applicando al documento elettronico un timbro o una firma digitale.

Una firma digitale può avere svariati aspetti: un nome scritto a mano, un logo, un'immagine oppure un semplice testo esplicativo della firma. A differenza di una comune firma analogica, nella firma digitale rimangono salvate informazioni riguardanti lo status di chi firma e le caratteristiche del documento. La firma sul documento è semplicemente un segno grafico, in realtà le informazioni si trovano altrove.

La tecnologia su cui si basano le misure di protezione è costituita da algoritmi di crittografia e di hashing sicuro.

Gli algoritmi di crittografia sono algoritmi matematici in grado di trasformare (cifrare) reversibilmente un insieme di dati, ad esempio un documento, in modo da renderli intelligibili. Affinché questi algoritmi siano di qualche utilità pratica occorre che soddisfino alcune condizioni fondamentali.

- La cifratura e la decifratura deve avvenire in funzione di una variabile detta chiave e costituita da una sequenza di bit di lunghezza variabile in funzione dell'algoritmo e del livello di sicurezza che si desidera ottenere.
- Le operazioni di cifratura e decifratura sono relativamente semplici nel caso in cui si conosca la chiave, in caso contrario risultano laboriose al punto da risultare praticamente inattuabili.
- Risulta egualmente laborioso dedurre la chiave con cui è stato cifrato un documento confrontandolo con la sua versione in chiaro.

La lunghezza delle chiavi, quando non fissata dal particolare algoritmo di crittografia, può tipicamente assumere un insieme di valori che varia in funzione

²⁶ GIOVANNI PASCUZZI, ROBERTO CASO, *op. cit.*, p. 314.

dell'algoritmo stesso e degli standard applicabili. La lunghezza effettivamente scelta per le chiavi da utilizzare nell'ambito di una specifica applicazione è sempre il risultato di un compromesso fra esigenze di sicurezza e potenze dei calcolatori a disposizione. Al crescere della dimensione delle chiavi, infatti, aumenta la sicurezza (intesa come difficoltà di decifrare le informazioni crittografate), ma anche la potenza di elaborazione necessaria per contenere i tempi delle operazioni di cifratura entro limiti accettabili²⁷.

Gli algoritmi di crittografia possono essere di due tipi: simmetrici o "a chiave privata" e asimmetrici o "a chiave pubblica".

I primi usano un'unica chiave per cifrare e decifrare, per questo motivo si prestano allo scambio di informazioni tra due soggetti, ma non alla comunicazione di informazioni a soggetti sconosciuti in un ambiente come il web, dal momento che sia chi effettua la cifratura che chi effettua la decifratura deve conoscere la chiave.

Un esempio di algoritmo simmetrico è il DES (Data Encryption Standard), le cui chiavi possono avere una lunghezza che varia tra i 56 ed i 112 bit.

Gli algoritmi asimmetrici, invece, utilizzano due diverse chiavi per cifrare e decifrare i documenti. Una chiave è detta pubblica, mentre l'altra è detta privata in quanto rimane di conoscenza solo del creatore della firma digitale. Nonostante le due chiavi vengano generate da un unico algoritmo, non è possibile da una risalire all'altra. L'algoritmo RSA, così chiamato perché creato da Rivest, Shamir e Adleman, è quello in questo momento più diffuso. Questo tipo di algoritmo genera coppie di chiavi da 512 o da 1024 bit.

Gli algoritmi asimmetrici, poiché sono più elaborati e quindi più lunghi da elaborare, solitamente vengono accostati ad algoritmi simmetrici e di hashing sicuro.

Gli algoritmi di hashing sicuro non servono a cifrare i documenti, ma a verificare che questi non abbiano subito manipolazioni. Questo è possibile in quanto l'algoritmo di hashing produce una sequenza di bit di lunghezza fissa strettamente connessa al documento, chiamata Digest. Mettendo a confronto

²⁷ VITTORIO CARACCIOLO, "La firma digitale",
<[http://db.formez.it/FontiNor.nsf/d5be50a8ff723448c1256a95005855f0/4F8FD8A4EF9E3076C1256C5A002E37FC/\\$file/firma_digitale.pdf](http://db.formez.it/FontiNor.nsf/d5be50a8ff723448c1256a95005855f0/4F8FD8A4EF9E3076C1256C5A002E37FC/$file/firma_digitale.pdf)>.

due digest prodotti dallo stesso documento si può verificare l'integrità del documento stesso; se i due digest risulteranno identici significherà che questi non sono stati alterati. L'algoritmo solitamente usato si chiama SHA, Secure Hash Algorithm.

Per il momento le immagini del progetto non sono protette da nessuna misura a livello tecnologico, ma si procederà al più presto a "timbrare" virtualmente tutti i documenti.

CAPITOLO 6: Sostenibilità e preservazione

LA SOSTENIBILITÀ

Una volta finiti i fondi per un progetto di digitalizzazione viene spontaneo chiedersi come saranno mantenute le risorse appena create. A questo quesito fondamentale si deve trovare una soluzione all'inizio del progetto. Una risorsa sostenibile è sostanzialmente quella che continuerà ad esistere a lungo dopo che i fondi iniziali per il progetto saranno finiti¹. La pianificazione per la sostenibilità di una risorsa fa parte della "strategia d'uscita" del progetto. La strategia d'uscita è il piano che descrive le opzioni di un progetto per il futuro dopo la fine dello stesso e deve essere naturalmente specifica per ogni tipo di progetto, ma la sostenibilità è il fattore chiave della strategia. Se una risorsa digitale non può essere sostenuta, tutto lo sforzo, il lavoro, il tempo ed il denaro per crearlo saranno stati investiti inutilmente, ecco perché la sostenibilità deve essere pianificata prima possibile.

Per fare un esempio il progetto "Archivo General de Indias" dal 1986 al 1992 ha richiesto complessivamente l'investimento di 6,6 milioni di dollari per la sua creazione, mentre nei due anni successivi, per il mantenimento del sistema, ha richiesto altri due milioni di dollari. Ha però permesso di ridurre del 31% il rischio di deterioramento degli originali dovuto all'utilizzo e del 38% il rischio di deterioramento causato dalle fotocopie.²

A proposito di sostenibilità di una risorsa il TASI propone i seguenti approcci³:

1. Manutenzione continua: si deve continuare a badare alla risorsa ed a mantenerla, ma ciò non significa che essa non debba essere sviluppata in termini di nuovi contenuti o caratteristiche aggiunte.

¹ TECHNICAL ADVISORY SERVICE FOR IMAGES "Sustainability of Digital Resources" in *Technical Advisory Service for Images – Advice Paper*
<<http://www.tasi.ac.uk/advice/creating/pdf/sust.pdf>>.

² PEDRO GONZALES GARCIA, "General System Architecture"
<<http://www.clir.org/pubs/reports/gonzales/architecture.html>>.

³ TECHNICAL ADVISORY SERVICE FOR IMAGES "Sustainability of Digital Resources" in *Technical Advisory Service for Images – Advice Paper*
<<http://www.tasi.ac.uk/advice/creating/pdf/sust.pdf>>.

2. Sviluppo continuo: la risorsa continua a crescere e/o a svilupparsi. Questo significherà costruire un lavoro su un progetto esistente e, in alcuni casi, fare la transizione da un progetto ad un servizio o ad un programma. Lo sviluppo richiederà ovviamente un ulteriore consistente finanziamento.
3. Integrazione: l'integrazione del materiale in una collezione esistente non coinvolgerà necessariamente ulteriori investimenti, ma richiederà una pianificazione in termini di interoperabilità ed uso degli standard.
4. Cambiamento delle finalità: l'interoperabilità è fondamentale se si pianifica di riutilizzare il materiale per un'altra risorsa. Una nuova finalità richiede modificazioni basate sul tipo di file, sulla dimensione dei file, sulla risoluzione o sulla modalità del colore da usare con un supporto diverso, o potrebbe consistere in un nuovo lavoro sui metadati per andare incontro ai bisogni di una nuova utenza.

Qualunque approccio sia scelto, il TASI identifica tre fattori chiave da considerare quando si pianifica la sostenibilità⁴:

1. Si devono innanzitutto esaminare i passi pratici che aiutino una risorsa ad essere sostenibile. I seguenti provvedimenti pratici possono aumentare le chance della sostenibilità:

- Design flessibile. Per le future migrazioni, le risorse dovrebbero essere pensate in modo che i singoli componenti, come i digital objects o gli elementi di presentazione, siano distinti e divisibili, così che ogni componente possa essere aggiornato, alterato o rimosso senza interferire con le altre parti del sistema.
- Approccio flessibile. I project manager e le altre persone coinvolte nelle decisioni dovrebbero mantenere un approccio flessibile per tutta la durata del progetto, senza perderne di vista gli scopi.
- Utilizzo di standard appropriati. Per assicurare l'interoperabilità e la longevità, dovrebbero essere usati formati, metadati, supporti per la conservazione e un sistema di ricerca standard. L'uso di standard in queste aree migliorerà la

⁴ TECHNICAL ADVISORY SERVICE FOR IMAGES "Sustainability of Digital Resources" in *Technical Advisory Service for Images – Advice Paper*
<<http://www.tasi.ac.uk/advice/creating/pdf/sust.pdf>>.

probabilità che la collezione o i suoi componenti vengano utilizzati e faciliterà il loro inserimento all'interno di altri progetti.

- Garanzia di qualità e sua valutazione. La garanzia di qualità ed un'ininterrotta valutazione per tutto il progetto assicurerà che la risorsa finita sia pertinente e valida per i suoi utenti, ed aiuterà ad identificare futuri sviluppi o nuove opportunità.
- Documentazione. Una diligente documentazione di tutti gli aspetti del progetto (sistemi e processi, metodologia del workflow, standard, schemi, formati, altri dettagli tecnici) è fondamentale per la sostenibilità di una risorsa oltre la durata del progetto.

2. Si deve poi capire che una risorsa sostenibile richiederà un crescente livello di manutenzione una volta che il progetto sarà finito. Il livello di manutenzione varierà enormemente in base all'approccio alla sostenibilità di ogni singolo progetto ed al tipo di pianificazione dello stesso. I tipi di mantenimento richiesti possono essere:

- Disponibilità, cioè sicurezza che la risorsa sia disponibile a tutti coloro che ne hanno bisogno quando la desiderano.
- Procedure di sicurezza, ovvero autenticazione dell'utente, gestione dei privilegi d'access, protezione dai virus, back-up del sistema e conservazione di copie d'archivio.
- Gestione dei diritti, intesa come sistema per controllare il corretto uso del patrimonio digitale.
- Tutela del contenuto, aggiornandolo con nuovo materiale, rivedendo o eliminando materiale esistente, aggiornando i metadati dei record esistenti, correggendo gli errori e mantenendo la risorsa dinamica e rilevante.
- Previdenza, in quanto il sistema deve stare al passo con le tecnologie emergenti, deve essere compatibile con i browser più recenti, conformarsi alle linee guida sull'accessibilità, essere rispondente ai bisogni dell'utenza quando cambiano.
- Amministrazione, per la quale è necessario uno staff e delle apparecchiature. Se un progetto può contare su uno staff esperto, ci sarà una transizione senza problemi dal progetto alla sua manutenzione.

3. Infine si devono esaminare le opzioni disponibili per finanziare la sostenibilità. Quest'ultimo fattore avrà, chiaramente, la maggiore influenza e determinerà in definitiva la sostenibilità. Per assicurarsi i fondi per mantenere il materiale disponibile oltre la durata del progetto, si deve considerare una combinazione delle seguenti possibilità:

- a. Fondi essenziali. A seconda del tipo di progetto e del valore percepito all'interno dell'istituzione che lo ospita, sarà possibile assicurarsi i fondi essenziali. Dato che le risorse digitali acquistano un profilo sempre più alto, dovrebbe diventare più semplice farle apparire come patrimonio istituzionale degno di un investimento a lungo termine.
- b. Nuovi fondi. La fonte del finanziamento iniziale del progetto potrebbe concedere ulteriori fondi o, in alternativa, si possono cercare nuovi finanziatori e sponsor.
- c. Vendita online. Si potrebbe prendere in considerazione la possibilità di recuperare le spese direttamente dal contenuto digitale ponendo diritti di riproduzione o vendendo copie fisiche all'utenza. Implementare una strategia di questo tipo potrebbe richiedere una sistema di gestione delle vendite e degli ordini, una tassazione del cliente, il pagamento delle royalty a parti terze. Questo processo può essere però molto complicato e dispendioso da intraprendere e non assicura un guadagno dalle vendite, a meno che la collezione non sia particolarmente adatta allo sfruttamento commerciale.
- d. Affidamento a parti terze. Si può decidere di affidarsi a parti terze in vari modi:
 - Abbonamento, per il quale si chiede agli utenti di pagare per ricevere un servizio aggiuntivo. É una situazione difficile da realizzare se il servizio è stato in precedenza gratuito, ma se si riesce ad offrire un valore aggiunto è un buon metodo per avere introiti.
 - Sponsorizzazione, che può essere di due tipi: uno o due sponsor principali, che pagano per il loro nome o i loro prodotti sul sito oppure un numero più ampio di singoli sponsor che pagano un piccolo contributo per le spese correnti. La prima soluzione può presentare alcuni problemi: potrebbe

essere difficile trovare uno sponsor adatto o una combinazione di sponsor disponibili a condividere lo spazio. Gli sponsor commerciali potrebbero fare specifiche richieste sulla quantità di spazio o su quanto deve essere visibile la loro pubblicità. La seconda opzione richiederà una grande quantità di lavoro promozionale per trovare singole persone o compagnie disponibili a sostenere il progetto e richiederà molti sforzi anno dopo anno.

- Pubblicità, in quanto si potrebbe far pagare ad altre organizzazioni una tariffa per pubblicizzare le loro merci. Tale soluzione, molto comune nei siti commerciali, potrebbe essere utilizzata per siti non commerciali a patto che essi abbiano una grande quantità di utenti.

Apparentemente sembrano molteplici i metodi per la raccolta di fondi per il sostegno di una risorsa digitale, ma ognuno di questi deve essere considerato con molta cautela. Nel caso delle vendite, della sponsorizzazione e della pubblicità si possono creare conflitti con la natura no-profit dell'istituzione di appartenenza e con gli utenti. Ogni incursione nel settore commerciale deve essere condotta assennatamente, rendendo chiaro a tutti (utenti, istituzioni di appartenenza ed altri finanziatori) che ogni guadagno sarà sfruttato per continuare ad offrire il servizio.

PRESERVAZIONE

Nella pianificazione di una "strategia d'uscita" per il progetto bisogna considerare anche l'aspetto della preservazione che può rientrare nel discorso della sostenibilità, ma che per la sua importanza merita di essere trattato a parte.

Nell'introduzione al volume "L'eclisse delle memorie", già citato in precedenza, Tullio Gregory scrive:

Uno dei tanti paradossi che sottendono le nostre società e del quale ci si rende ogni giorno più consapevoli, è costituito dalla compresenza di due dinamiche in opposta tensione: da un lato lo smisurato aumento di documenti prodotti dalle strutture culturali, economiche, industriali, istituzionali, dall'altro il rapido processo di decomposizione e di perdita dei documenti stessi.⁵

⁵ TULLIO GREGORY, MARCELLO MORELLI, *op. cit.*, p. XI.

Ed in particolare ai supporti informatici, soggetti a processi di rapida obsolescenza, attribuisce questa paradossalità:

... al gigantismo delle memorie corrisponde la perdita rapidissima sia dei sistemi di codifica, sia degli strumenti tecnici di decodifica, quindi di lettura.⁶

La conservazione è stata e continua ad essere un compito fondamentale per la biblioteca tradizionale; per la biblioteca digitale diventa un problema maggiore in quanto non esistono enti deputati a questo compito, dal momento che i supporti sono più volatili (i supporti cartacei e membranacei hanno dimostrato di avere una durata superiore ai 500 – 1000 anni) e che si potrebbe verificare la possibilità che il documento diventi illeggibile se non si conserva anche il software.

La preservazione ha una componente rilevante nella gestione, è pertanto importante per il successo di ogni collezione digitale impostare una strategia sin dall'inizio del progetto di creazione di una digital library.

Definizione

In ambito digitale si parla di preservazione in quanto il termine ingloba in sé una concezione più ampia rispetto alla conservazione.

La preservazione è quella funzione della biblioteca digitale che comprende le attività di memorizzazione dell'informazione su supporti digitali, il mantenimento delle collezioni digitali, l'accesso permanente e la distribuzione dei documenti digitali.⁷

La preservazione consiste quindi in una serie di attività che assicurino la durata, la sicurezza e l'integrità intellettuale dell'informazione.

Con la durata si intende garantire l'accesso istantaneo e perpetuo, con la sicurezza si identifica il processo che deve prevenire la modifica dei dati memorizzati in formato digitale, mentre l'integrità sta ad indicare il processo che deve prevenire danni ai documenti digitali da incidenti esterni.

⁶ TULLIO GREGORY, MARCELLO MORELLI, *op. cit.*, p. XII.

⁷ ALBERTO SALARELLI, ANNAMARIA TAMMARO, *op. cit.*, p. 165.

Oya Rieger semplifica la sua definizione puntando sulla sua finalità, preservare il digitale significa:

mantenere le collezioni digitali di immagini in una forma usabile e interpretabile per lungo tempo⁸

L'espressione "per lungo tempo" non significa per sempre, ma sta ad indicare che si deve riuscire a mantenerne la memoria per un periodo di valore continuo.

La memoria elettronica viene spesso paragonata alla tradizione orale; i due tipi di memoria sono simili per due aspetti fondamentali. Scrive Luciana Duranti:

In primo luogo, tutti i loro contenuti sono conservati nella stessa forma estrinseca, una forma che non permette un'immediata e facile identificazione della natura e dello scopo del messaggio conservato e/o trasmesso [...]. In secondo luogo, la conservazione di entrambe dipende da uno sforzo attivo, da un passaggio sistematico dei loro contenuti da un supporto o da un contenitore ad un altro, da una periodica trasformazione delle informazioni in formati intelligibili e contemporanei, dall'esistenza di garanzie di affidabilità contestuali e di prova della loro autenticità esterna ad esse, e da una continua eliminazione del ridondante e del superfluo.⁹

Patricia Battin, alla 59. Conferenza dell'IFLA, a Barcellona nel 1993, ha addirittura sostenuto che:

[...] Nel mondo digitale, preservazione è accesso e accesso è preservazione.¹⁰

Obiettivi

"La condizione per il recupero delle informazioni conservate è duplice: la conservazione del testo e la conoscenza del meccanismo di accesso all'informazione"¹¹.

Perché questa duplice condizione si avveri, la preservazione si pone come obiettivi:

⁸ ANNE KENNEY, OYA RIEGER, *op. cit.*, p. 135.

⁹ LUCIANA DURANTI, "La definizione di memoria elettronica", in TULLIO GREGORY, MARCELLO MORELLI, *op. cit.*, p. 148.

¹⁰ HANS RUTIMAN, "La conservazione del nostro patrimonio culturale. Un interesse di tutti" in TULLIO GREGORY, MARCELLO MORELLI, *op. cit.*, p. 7.

¹¹ HEINZ ZEMANEK, "L'informazione è sorpresa" in TULLIO GREGORY, MARCELLO MORELLI, *op. cit.*, p. 219.

- L'identità dei bit per assicurare l'integrità del contenuto attraverso
 - il monitoraggio della corruzione per la stabilità e l'autenticità dei dati,
 - la protezione del contenuto da alterazioni non documentate,
 - la sicurezza dei dati da usi non autorizzati fornendo la stabilità dei media.
- Il contesto tecnologico per mantenere le interazioni fra elementi di un ambiente digitale più vasto
 - preservando il contesto, assicurando cioè che il file dei metadati e gli script che collegano questi file alle immagini siano integri,
 - mantenendo l'integrità dei collegamenti,
 - monitorando la creazione dinamica dei documenti (conversione on the fly o aggiornamenti periodici dei metadati).
- La provenienza per mantenere un tracciato dell'origine e della storia del contenuto, per esempio descrivendo la fonte del contenuto digitale e le alterazione che possono essere intervenute fin dalla sua creazione, come aggiornamenti di immagini o metadati o cambiamenti sul supporto di conservazione.
- I riferimenti e l'usabilità per assicurare che gli utenti possano facilmente collocare, cercare e usare la collezione digitale; per esempio creando degli identificatori univoci come i PURL.

Problemi posti dal digitale

Come precedentemente affermato, in ambiente digitale, non si parla di conservazione del materiale, ma di preservazione.

La preservazione gestisce l'intero ciclo di vita del documento digitale, con un'enfasi sull'accesso di lunga durata all'informazione digitale, conservata nella sua integrità.¹²

Si pensa alla memoria elettronica come formata da due strutture distinte e separate, una struttura fisica ed una struttura logica. In realtà queste due strutture sono strettamente collegate e dipendenti l'una dall'altra. Se il supporto

¹² ALBERTO SALARELLI, ANNAMARIA TAMMARO, *op. cit.*, p. 165.

fisico non viene protetto e custodito nel migliore dei modi l'obiettivo ultimo non può essere raggiunto.

La salvaguardia delle raccolte, quali che siano i documenti in esse presenti, deve tradursi in un'attenzione costante alle condizioni di tali raccolte, alla qualità intrinseca degli oggetti che si prevedono di conservare, al loro stato fisico, alla loro età, alle specifiche condizioni di mantenimento e di tutela.¹³

Tutti i supporti che conservano informazioni in formato digitale sono soggetti al decadimento fisico, per questo l'integrità delle informazioni deve essere innanzitutto preservata a questo livello. L'alterazione dei supporti ottici (come i DVD usati per il salvataggio delle immagini master) può diventare una grave causa di perdita dei dati: anche un semplice graffio infatti può compromettere la corretta lettura da parte del laser rendendo il supporto ottico difficilmente riutilizzabile. Al momento non esistono supporti digitali più durevoli e affidabili di quelli analogici, ma si deve essere fiduciosi in un continuo sviluppo e miglioramento di tali supporti grazie alle nuove tecnologie. Nella sua documentazione il TASI sottolinea come i supporti su cui vengono salvate le informazioni si sono modificati negli ultimi decenni come mostra il diagramma di seguito riportato:

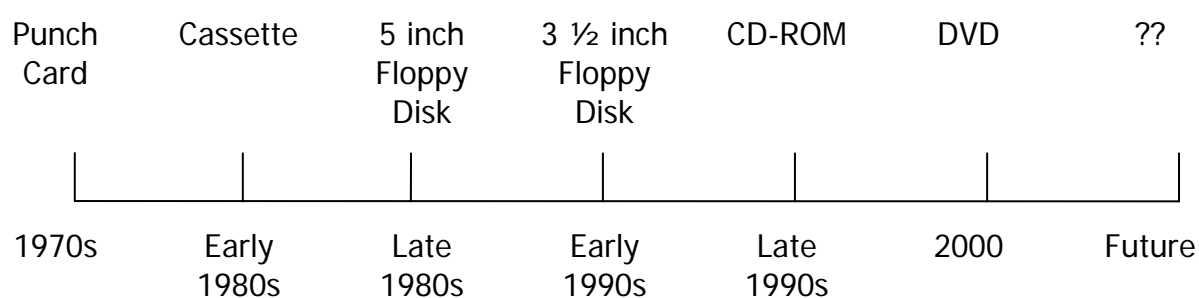


Fig1: la linea del tempo dei più comuni supporti di salvataggio¹⁴.

I dischi ottici sono formati da una serie di strati sovrapposti; il rivestimento, riflettente, è solitamente in alluminio, mentre la parte più interna è realizzata in policarbonato.

¹³ MICHELE SANTORO, "Dall'analogico al digitale: la conservazione dei supporti non cartacei", *Biblioteche oggi*, 3 (19) 2001: 88 – 100.

¹⁴ TECHNICAL ADVISORY SERVICE FOR IMAGES "An introduction to Digital Preservation" in *Technical Advisory Service for image , Advice Paper* <<http://www.tasi.ac.uk/advice/creating/pdf/digipres.pdf>>.

poiché i dischi sono costituiti sia di polimeri che di metalli, essi presentano tutti i problemi di degrado tipici di queste sostanze: i metalli infatti tendono alla corrosione e alla delaminazione, mentre i polimeri vanno incontro a deformazione e degrado. [...] In particolare è lo strato metallico a risultare il più debole, in quanto è più sottoposto alla corrosione: alcune leghe metalliche, infatti, in condizioni ambientali inadeguate, possono andare incontro a degrado e modificarne le proprietà. In sintesi le cause in grado di determinare una perdita di informazione contenuta nei supporti ottici sono dovute a influenze ambientali, quali l'inquinamento atmosferico, la temperatura e l'umidità, l'esposizione alla luce, ai raggi ultravioletti e ai campi magnetici parassiti; o a fattori manuali, ad esempio urti provocati dalla caduta dei dischi, graffi, e abrasioni, presenza di residui oleosi e altre impurità; oppure a fattori meccanici, ad esempio problemi di degrado dell'hardware provocati da un disallineamento del raggio laser e così via. Come tutti i supporti anche quelli ottici temono fortemente l'umidità e le variazioni di temperatura: questi due elementi pertanto devono fluttuare il meno possibile, e gli ambienti per la consultazione devono presentare le stesse condizioni atmosferiche dei magazzini; è dunque opportuno mantenere la temperatura tra i 5° e il 20° C e il tasso di umidità relativa tra il 20 ed il 50%

I documenti digitali, non sono solo più fragili dei loro corrispondenti analogici, ma più facilmente in grado di diventare fisicamente e logicamente inaccessibili: difatti il deterioramento dei supporti, la volatilità dei meccanismi di presentazione dei dati e la perdita di funzionalità dei sistemi di accesso sono tutti fattori che contribuiscono a rendere irraggiungibili un numero assai elevato di documenti. I documenti elettronici inoltre non possono essere utilizzati senza l'ausilio di particolari attrezzature, per cui un discorso sulla loro conservazione richiede la conoscenza accurata non solo dei supporti fisici destinati a contenerli, ma anche delle apparecchiature che permettono di crearli e utilizzarli e dei programmi senza dei quali il loro impiego non sarebbe possibile.¹⁵

Di ciò non si può essere certi perché, a detta di Pierluigi Ridolfi, "si possono fare ipotesi e simulazioni, ma non esistono evidenze di lungo periodo. [...] è garantita meglio la preservazione di un testo scritto su un'antica pergamena – perché il supporto resiste e resisterà nei secoli e perché si troverà sempre qualcuno in grado di leggere le antiche grafie - piuttosto che la sua trascrizione elettronica."¹⁶

¹⁵ MICHELE SANTORO, "Dall'analogico al digitale: la conservazione dei supporti non cartacei", *Biblioteche oggi*, 3 (19) 2001: 88 – 100..

¹⁶ PIERLUIGI RIDOLFI, "Tecnologia e memoria: l'evoluzione dei calcolatori e la conservazione dell'informazione" in TULLIO GREGORY, MARCELLO MORELLI, *op. cit.*, p. 251.

Il problema fondamentale, secondo Mirjam Foot, è che "questi nuovi mezzi non sono mai stati progettati per essere durevoli, di conseguenza sono molto meno stabili dei materiali convenzionali"¹⁷.

La studiosa propone tre principi base per conservare i supporti fisici elettronici:

1. archivarli separatamente da carta ed altri supporti a base di cellulosa,
2. creare più copie da conservare separatamente,
3. verificare regolarmente, tramite test, che le informazioni siano accessibili ed integre.

Un esempio significativo di quanto detto viene dalla storia dell'Archivo General de Indias, uno dei primi progetti di biblioteca digitale. In più di dieci anni sono state necessarie due migrazioni vere e proprie e quattro passaggi da un supporto all'altro. Gran parte del lavoro è derivato dal fatto che quando il progetto è partito non esistevano ancora o non erano ancora diffusi gli standard ed i supporti, oggi largamente usati. Tale esperienza dà un'idea di quali e quante problematiche si devono affrontare.¹⁸

Componenti di una politica di preservazione del digitale

Una strategia di preservazione presenta due livelli, una strategia di preservazione tecnica ed una strategia organizzativa. La prima si occuperà di stabilire come le informazioni digitali saranno rese usabili e accessibili attraverso i cambiamenti che possono intervenire. La seconda assicurerà che il progetto abbia sempre i fondi, le capacità ed uno staff preparato così che le strategie a livello tecnico possano essere messe in atto.

Quale sia la migliore strategia tecnica per la preservazione è un argomento molto dibattuto nella comunità biblioteconomica e non ha ancora avuto una risposta. Sicuramente una strategia solo a livello tecnico non è sufficiente da sola ad assicurare l'accessibilità e la leggibilità a lungo termine. È indispensabile

¹⁷ MIRJAM FOOT, "Aspetti della conservazione di massa", in TULLIO GREGORY, MARCELLO MORELLI, *op. cit.*, p. 90.

¹⁸ PEDRO GONZALES GARCIA, "Long Term Projects and Changing Methods. The case of the Archivo General de Indias". <<http://www.cs.uu.nl/events/dech1999/cv/gonzalez.htm>>.

che venga sviluppata anche una strategia organizzativa che prenda in considerazione¹⁹:

- una strategia di preservazione a livello tecnico,
- un budget per rendere operativa la preservazione della tecnologia,
- un programma di formazione per permettere a tutto lo staff del progetto di capire le problematiche legate alla preservazione digitale e alla sicura manipolazione delle informazioni,
- un gruppo di gestione o guida per sorvegliare tutte le problematiche rilevanti per la sicurezza ed il successo della preservazione.

Le componenti chiave di una politica di preservazione, secondo la Rieger²⁰, sono quattro:

- un'infrastruttura organizzativa
- una politica per la selezione e la conversione
- delle strategie di preservazione
- una previsione tecnologica

Infrastruttura organizzativa

Bisogna identificare le capacità ed i requisiti dello staff necessari per implementare le strategie di preservazione e fornirgli una formazione ed opportunità di sviluppo professionale.

Abby Smith sostiene che:

a fully staffed preservation department will reflect the hybrid nature of a library's holdings, with staff having, or having access to, the necessary expertise to handle a variety of media.²¹

È importante pertanto prevedere un impegno finanziario per lo staff, la sua formazione e l'aggiornamento dell'hardware e del software, anche se a priori è difficile prevedere il budget necessario e trovare i fondi.

¹⁹ TECHNICAL ADVISORY SERVICE FOR IMAGES, "Establishing a Digital Preservation Strategy " in *Technical Advisory Service for image , Advice Paper* <<http://www.tasi.ac.uk/advice/creating/pdf/digipres2.pdf>>.

²⁰ ANNE KENNEY, OYA RIEGER, *op. cit.*, p. 136 – 146.

²¹ ABBY SMITH, "Preservation in the future tense", in *CLIR issues*, 3 (1998). <<http://www.clir.org/pubs/issues/issues03.html>>.

Strategie di preservazione

Nella scelta del formato²² dei file e del supporto di memorizzazione occorre valutare la flessibilità nello spostamento dei dati a un nuovo supporto e/o un differente supporto di memorizzazione in un futuro osservabile. La conformità agli standard è un ragionevole indicatore che un particolare formato o supporto avrà valenza nel futuro. La migrazione da un formato ad un altro dovrebbe permettere di migrare da un formato senza perdita a uno con perdita.

Ogni scelta di formato probabilmente diventerà obsoleta in meno di 5 anni, per questo i supporti di memorizzazione, per prevenire la perdita dei dati, dovrebbero essere periodicamente aggiornati, al massimo ogni due o tre anni. Questo significa fare una copia di tutti i dati su un nuovo supporto. A proposito di migrazione Michael Cain, in un articolo intitolato "Being a library of record in a digital age" racconta la sua personale esperienza:

Scrisi in formato digitale una novella nel 1985. Quando finalmente la pubblicai nel 2000, ero passato attraverso tre piattaforme hardware, tre tipi di dischetti, cinque sistemi operativi, tre word processor e molteplici versioni di questi software. La mia esperienza illustra le sfide della migrazione dei dati. Quando le tecnologie maturarono e furono standardizzate, le cose diventarono più semplici, ma la migrazione non era mai perfetta, e solo perché ero l'autore e conoscevo il testo tanto bene ero in grado di mantenere la sua integrità. La sola migrazione di un item ad un altro supporto di conservazione non sarà sempre sufficiente. Le operazioni di preservazione digitale dovranno confrontarsi con tutte queste problematiche.²³

Le strategie più frequentemente messe in atto per la preservazione sono il refreshing, la preservazione della tecnologia, la migrazione, l'emulazione dell'ambiente hardware e software necessario.

Il **refreshing** consiste nella copia dell'informazione digitale su supporti più nuovi prima che i vecchi si deteriorino fino a compromettere la lettura ed il riversamento dei dati. Anche se esso non può essere considerato una vera e propria strategia di preservazione, è indubbiamente un'operazione utile

²² MINERVA WORKING GROUP, "Handbook of Good Practice",
<http://www.minervaeurope.org/structure/workinggroups/goodpract/document/goodpractices1_3.pdf>.

²³ MICHAEL CAIN, "Being a library of record in a digital age" in *Journal of Academic Librarianship*, 29 (6) 2003: 405-410.

- come cura preventiva per controllare l'invecchiamento e la decadenza che rendono i supporti poco affidabili e facilmente danneggiabili,
- per evitare l'obsolescenza dei supporti,
- per amalgamare le collezioni in una migliore integrazione e gestione,
- per avere il sostegno di tecnologie di salvataggio più efficienti e fidate che offrono capacità più elevate, migliore accesso e prezzi più bassi.

Naturalmente questa scelta può portare anche degli svantaggi²⁴:

- Non è prevedibile una programmazione in quanto essa dipende da diversi fattori come la frequenza d'uso e le condizioni di salvataggio, che intralciano una pianificazione a lungo termine;
- Si potrebbero verificare problemi di copyright se l'uso è limitato ad un preciso supporto di salvataggio per il controllo dell'accesso;
- Il trasferimento ad un nuovo supporto può cambiare l'ordine dei bit anche se non intenzionalmente;
- È complicato mantenere ogni informazione etichettata associata al supporto originario, poiché i supporti diventano più capienti.

La **preservazione della tecnologia** si fonda sulla preservazione dei dati oltre che dell'ambiente tecnologico hardware e software da cui dipendono. Per alcuni documenti digitali, almeno nel breve periodo, questa può essere la migliore soluzione poiché:

- assicura l'accessibilità del materiale attraverso gli strumenti che riescono a leggerlo;
- può risultare una misura di fortuna per quelle risorse digitali di valore che non possono essere convertite in formati indipendenti dall'hardware e/o dal software e fatte migrare altrove;
- preserva i comportamenti e la forma originale degli oggetti digitali.

Nel lungo periodo questa soluzione potrebbe diventare impraticabile in quanto:

- le esigenze di spazio, i costi di mantenimento (bug del software, risoluzione dei problemi, sostituzione delle componenti hardware), il

²⁴ ANNE KENNEY, OYA RIEGER, *op. cit.*, p. 146.

naturale invecchiamento del computer ed il mantenimento delle competenze dello staff diventano troppo dispendiosi;

- non è graduale e non può essere usata per qualsiasi tipo di collezione.

Nell'**emulazione** dell'ambiente hardware e software necessario alla conservazione del supporto di informazione l'aspetto e la funzionalità di una risorsa digitale è ricreato con l'emulazione delle caratteristiche dell'ambiente tecnologico superato. Questa strategia si basa sul sistema di metadati a corredo del documento digitale che descrivono l'ambiente tecnologico. L'emulazione è l'opzione "just in time" nell'ambito della preservazione.

La scelta di questa strategia è giustificata dalla possibilità:

- di supportare l'intento originale del documento digitale ricreando la sua funzionalità originale, il suo aspetto, la sua percezione;
- di sviluppare una strategia di preservazione graduale ed espandibile per altre collezioni simili;
- di avere una strategia di back-up per la migrazione.

Gli svantaggi possono però essere molteplici.

- L'emulazione potrebbe essere tecnicamente poco fattibile, richiedere un salto nel buio per quanto riguarda le tecnologie future e le capacità dello staff, risultare molto costosa senza offrire garanzie sulla riuscita.
- Essa presume che gli utenti in futuro capiranno come usare vecchi sistemi.
- Accresce potenzialmente le istanze di proprietà intellettuale riguardo all'acquisto ed al mantenimento di informazioni proprietarie necessarie per emulare hardware e software.
- Si focalizza maggiormente sulla preservazione delle funzionalità più che sull'integrità e l'autenticità dei dati.
- La risoluzione dei software bug e dei virus è molto difficoltosa.
- Il successo dipende dalla creazione e cattura dei metadati.
- Alcune componenti hardware, come RAM, schede video e audio, e i convertitori analogici sono difficili da emulare.

- Il rinnovo dei media è parte della strategia di emulazione con i suoi rischi.

La **migrazione** è il processo di trasferimento dell'informazione da un'ambientazione hardware e software a un'altra o da una generazione di computer ad una successiva²⁵. Questo processo comporta dei rischi di perdita dei dati per l'incompatibilità tra i due sistemi. Anche questa strategia offre vantaggi quali:

- l'aumento delle funzionalità del sistema man mano che la tecnologia si evolve,
- il salvataggio dei dati digitali che sono al limite dell'obsolescenza,
- l'aggiornamento di un sistema che può avere problemi o difetti,
- la creazione di un ambiente tecnico più stabile e sicuro,
- l'accrescimento della funzionalità della collezione ed il risparmio di tempo e denaro,
- il consolidamento delle collezioni nella stessa infrastruttura tecnologica.

Tra gli svantaggi si annoverano i seguenti.

- La pianificazione è difficoltosa poiché richiede una costante attenzione ai cambiamenti dei formati, del software di reperimento, del sistema operativo e del hardware.
- È necessario un lavoro intensivo, che richiede pianificazione ed analisi. Ogni ciclo può essere una nuova esperienza e difficilmente si riescono a rendere le tecniche di migrazione automatiche o semiautomatiche.
- Gli sforzi ed i rischi cambiano in base alla complessità del processo.
- Perdere un'opportunità può significare rendere non fattibile la migrazione dei file.
- Le ulteriori conversioni possono corrompere un documento digitale o modificare sostanzialmente il modo in cui era stato pensato dall'autore e dalla sua utenza.

²⁵ ANNE KENNEY, OYA RIEGER, *op. cit.*, p. 147 – 148.

- Le piccole perdite possono essere difficili da rilevare e valutare e sorge un problema quando la migrazione modifica il comportamento dell'immagine.
- L'assenza di specifiche chiare della fonte e dell'ambiente destinatario contribuisce in modo significativo all'aumento dei costi e ad una minore affidabilità.
- Ci potrebbero essere delle restrizioni legali.

Leggendo la vasta documentazione tecnica riguardo alla preservazione si ha l'idea che la metodologia che riscuote maggiori consensi è la migrazione, ma, come scrive Stefano Vitali, questa "cancella via via le tracce del contesto tecnologico all'interno del quale i documenti sono stati posti in essere , può comportare la completa scomparsa di quella capacità, che i documenti ed i reperti storici in genere hanno, di suggerire, attraverso la loro sembianza fisica, il senso dello scorrere del tempo."²⁶

Vitali propone come alternativa la strategia dell'emulazione che consente di riprodurre componenti significative dei contesti tecnologici in cui le fonti digitali hanno avuto origine. Le fonti storiche non trasmettono solo informazione verbale, anche la componente fisica e contestuale riveste ed ha sempre rivestito una notevole importanza nella loro interpretazione; in definitiva

c'è una dimensione del rapporto con le fonti che non si esaurisce nella quantità di informazioni che da esse è possibile ricavare, ma che investe la sua qualità profonda, il complesso delle sensazioni e dei meccanismi cognitivi – compresi quelli percettivi ed estetici – che esse sono in grado di evocare e mettere in moto. E questo è tanto più vero per fonti come quelle digitali, i cui stili comunicativi sono spesso caratterizzati da una stretta integrazione di testi, grafica, immagini e suoni, che sollecita un nuovo ordine dei sensi e richiede forme di intelligenza simultanea, caratterizzate dalla capacità di trattare nello stesso tempo più informazioni....²⁷

STRATEGIA DEL PROGETTO BENUSSI

La "strategia d'uscita" del progetto di valorizzazione del Fondo Benussi non è stata incentrata sulla preservazione a lungo termine, ma su uno sviluppo ed un

²⁶ STEFANO VITALI, *op. cit.*, p. 183.

²⁷ *Ibidem*.

ampliamento del progetto che potrebbe in gran parte rivoluzionare la struttura della Digital Library nel suo insieme. Venuti a conoscenza della curiosità e dell'interesse che il progetto ha suscitato nella comunità scientifica, dietro proposta di alcuni studiosi e grazie alle donazioni di altri fondi di importanti psicologi italiani della generazione di Benussi o posteriori, con l'appoggio della stessa università, si è decisa la costituzione di un centro interdipartimentale per la psicologia in collaborazione tra il Dipartimento di Psicologia ed il Dipartimento di studi giuridici dell'Università di Milano Bicocca.

Oltre all'acquisizione di nuovo materiale librario e archivistico, che andrà ad incrementare la collezione primaria, si è pensato di creare una collezione secondaria formata dal materiale citato nei documenti della collezione primaria e da eventuali contributi legati al materiale presente nel centro, che avrà anche una sezione dedicata agli studi sulla psicologia forense.

Per fare fronte a questo ampliamento del progetto è stato richiesto un nuovo finanziamento al Ministero dell'Università e della Ricerca, un finanziamento ad una fondazione privata e sono stati stanziati dei fondi da parte dei Dipartimenti che partecipano al Centro di documentazione. Si è anche presa in considerazione la possibilità di far pagare un abbonamento per accedere ai documenti, ma per il momento questa opzione è stata esclusa in quanto non si ha la sicurezza che le spese aggiuntive per la creazione di un sistema per la gestione delle transazioni economiche avrebbero il riscontro desiderato in termini di guadagni.

Nel frattempo comunque ci si è preoccupati di fare copia di tutti i DVD su cui erano salvate le immagini master. Tali DVD saranno conservati nei depositi della biblioteca, mentre i DVD originali rimarranno nei locali del centro di documentazione. Di tutte le immagini in formato jpg ed in formato PDF è stata fatta una copia di back-up. Oltre a ciò è già stata compiuta una prima migrazione dei metadati alla nuova versione di ALEPH 16, acquistata dalla biblioteca nel giugno 2004.

L'unica procedura che avrà una programmazione ben precisa sarà quella, a breve termine, del back-up dei dati sul server in base alla quale è prevista la copia di tutti i dati su nastro una volta al mese, a meno che non vengano effettuati degli aggiornamenti.

BIBLIOGRAFIA

"About the Iniziative" <<http://www.dublincore.org/aboutDCMI.htm>>.

ALEPH500 System Administration Guide, Release 16.1, 2004.

AMEN, KEOGH, WOLFF, "Digital Copyright: A tale of domestic discord, presented in three acts" in *Computers in Libraries*, 22(5) 2002.

MAXWELL ANDERSON, "Electronic Information and Digitization: Preservation and Security Challenges" in *Journal of Library Administration*, 38 (3/4) 2003, p. 163 – 168.

FRANCA ARDUINI, "Rinascimento virtuale" in *Biblioteche oggi*, 8 2002.

WILLIAMS ARMS, CHSISTOPHE BIANCHI et al, "An Architecture for Information in Digital Libraries" in *D-lib Magazine*, 3 (2) 1997.
<<http://www.dlib.org/dlib/february97/02arms.html>>.

WILLIAMS ARMS, "Uniforme Resource Name. A Progress Report" in *D-Lib Magazine*, 2 (2) 1996 <<http://www.dlib.org/dlib/february96/02arms.html>>.

ARUN SEN, "Metadata management: past, present and future", in *Decision Support System*, 37 (1) 2004: 151-173.

CATHERINE AYRE, ADRIENNE MUIR, "The right to preserve: the rights issues of digital preservation" in *D-Lib Magazine*; 10 (3) 2004.

LUCA BARDI, "Prende forma la digital library" in *Biblioteche oggi*, 12 1998.

NEIL BEAGRIE, DANIEL GREENSTEIN, "A strategic policy framework for creating and preserving digital collection",
<<http://ahds.ac.uk/manage/framework.html>>.

DAVID BEARMAN, "Reality and chimeras in the preservation of electronic records" in *D-Lib Magazine*; 5 (5)

VITTORIO BENUSSI, MAURO ANTONELLI (Hrsg), *Psychologische Schriften*, Amsterdam – New York, Rodopi, 2002.

TIM BERNERS LEE, "Metadata Architecture". January 1997.
<<http://www.w3.org/DesignIssues/Metadata.html>>.

CRISTINE BORGMAN, "What are Digital Libraries? Competing visions" in *Information Processing and Management* , 38(3) 1999: 227-243.

MICHAEL CAIN, "Being a library of record in a digital age" in *Journal of Academic Librarianship*; 29 (6) 2003: 405-410.

VITTORIO CARACCILO, "La firma digitale",
<[http://db.formez.it/FontiNor.nsf/d5be50a8ff723448c1256a95005855f0/4F8FD8A4EF9E3076C1256C5A002E37FC/\\$file/firma_digitale.pdf](http://db.formez.it/FontiNor.nsf/d5be50a8ff723448c1256a95005855f0/4F8FD8A4EF9E3076C1256C5A002E37FC/$file/firma_digitale.pdf)>

CENTRE COMMUN DE MICROELECTRONIQUE DE CROLLES, "Language Independent Metadata Browsing of European Resources"
<<http://dbs.cordis.lu>>.

FABIO CIOTTI, GINO RONCAGLIA, *Il mondo digitale. Introduzione ai nuovi media*, Roma – Bari, Laterza, 2000.

VALENTINA COMBA, "La qualità della comunicazione nella biblioteca digitale" in *Biblioteche oggi*, (9) 2000: 72 – 77.

COMMISSION OF PRESERVATION AND ACCESS – RESEARCH LIBRARY GROUP, "Preserving digital information. Report of the task force on the archiving of digital information", <<http://www.rlg.org/ArchTF/tfadi.index.html>>.

THOMAS CONNERTZ, Long-term digital archiving of digital documents: what efforts are being made in Germany?", in *Learned publishing*, 16 (7) 2003: 207 – 211.

"DCMI Metadata Terms" <<http://www.dublincore.org/documents/2004/06/14/dcmi-terms/>>.

"DCMI history" <<http://www.dublincore.org/about/history.htm>>.

DEAN, "Digital imaging and conservation: model guidelines" in *Library Trends*, 52 (1) 2003: 133-137.

WEIBEL DEKKERS, "State of the Dublin Core Metadata Initiative, April 2003" in *D-Lib Magazine*, 9 (4) 2003.

LORCAN DEMPSEY, ANNE MUMFORD *et al.* eLib Standards Guidelines – version 2.
<<http://www.ukoln.ac.uk/services/elib/papers/other/standards/version2.html>>.

ANTONELLA DE ROBBIO, "La biblioteca nel Web, il Web nella biblioteca" in *Bibliotime*, 2 (2) 1999. <<http://spbo.unibo.it/bibliotime/num-ii-2/derobbio.html>>.

ANTONELLA DE ROBBIO, "Metadati: parola chiave per l'accesso alla biblioteca ibrida"; Convegno La biblioteca ibrida, Milano, 14 – 15 marzo 2002.

ANTONELLA DE ROBBIO, "Vademecum on-the-fly per progetti di digitalizzazione" in E-LIS; 2003, <<http://eprints.rclis.org/archive/00000137>>

DIETMAR, WOLFRAM, XIE, HONG, "Traditional IR for web user: a context for general audience digital libraries" in *Information Processing & Management*, 38 2002: 627 – 648.

MARIELLA DI GIACOMO, MARK MARTINEZ, JEFF SCOTT, RICK LUCE "A fault-tolerant architecture for supporting large scale digital libraries" in *Issues in Science and Technology Librarianship*; (38) 2003. <<http://www.istl.org/03-summer/article3.html>>.

Digital Libraries: People, Knowledge, and Technology, 5th International Conference on Asian Digital Libraries, Singapore 2002, Berlin, Springer- Verlag, 2002.

DIGITAL LIBRARY FEDERATION – RESEARCH LIBRARIES GROUP, Guide to quality in Visual Resource Imaging, 2000. <<http://www.rlg.org>>.

MARTIN DILLON, "Metadata for Web Resources: How Metadata Works on the Web". <http://www.loc.gov/catdir/bibcontrol/dillon_paper.html>.

"Dublin Core Metadata Element Set, versione 1.1: traduzione italiana". <<http://www.iccu.sbn.it/dublinco.html>>.

ELECTRONIC TEXT CENTER, "Archival Digital Image Creation" <<http://etext.lib.virginia.edu/helpsheets/specscan.html>>.

PIERO FALCHETTA, "Guida breve alla digitalizzazione in biblioteca" in *Biblioteche oggi*, 18 (9) 2000: 52 – 67.

FINDLAY, "Future proof: ensuring the long-term accessibility of technology-dependent records" in *Records Management Journal*, 12 (3) 2002: 87-93.

FRANCOIS FLUCKIGER, *Understanding networked multimedia: applications and technology*, London- New York, Prentice Hall, 1995.

AMY FRIEDLANDER, DEANNA MARCUM, "A Keepers of the crumbling culture: what digital preservation can learn from library history" in *D-Lib Magazine*, 8 (5) 2003. <<http://www.dlib.org/dlib/may03/friedlander/05friedlander.html>>.

ANNA GALLUZZI, "Strumenti di valutazione per i servizi digitali", in *Biblioteche oggi*, (12) 2001: 6 – 13.

Giornata di studio sulla Digitalizzazione e Archiviazione delle immagini, ITIM, Milano, 23 marzo 2000 <<http://www.itim.mi.cnr.it>>.

HENRY GLADNEY, EDWARD FOX *et al*, *Digital Library: Gross Structure and Requirements: Report from a march 1994 Workshop*. <<http://www.csdl.tamu.edu/DL94/paper/fox.html>>.

PEDRO GONZALES GARCIA, "General System Architecture"
<<http://www.clir.org/pubs/reports/gonzales/architecture.html>>.

PEDRO GONZALES GARCIA, "Long Term Projects and Changing Methods. The case of the Archivo General de Indias".
<<http://www.cs.uu.nl/events/dech1999/cv/gonzalez.htm>>.

STEWART GRANGER, "Emulation as a digital preservation strategy" in *D-Lib Magazine*; 6 (10) 2000.

STEWART GRANGER, "Digital preservation and deep infrastructure" in *D-Lib Magazine*; 8 (2) 2002.

TULLIO GREGORY, MARCELLO MORELLI (a cura di), *L'eclisse delle memorie*, Roma – Bari, Laterza, 1994.

MARIEKE GUY, ANDY POWELL, MICHAEL DAY, "Improving the Quality of Metadata in Eprint Archives", in *Ariadne*, (38).
<<http://www.ariadne.ac.uk/issue38/guy>>.

FRITZ HEIDER, "Gestalt Theory: Early History and Reminiscences", in HENLE, JAYNES, SULLIVAN (eds.), *Historical Conceptions of Psychology*, New York, Springer, 1973.

TERRY HANSON, "The Access Catalogue Gateway to Resources" in *Ariadne*, (15). <<http://www.ariadne.ac.uk/issue15/hanson>>.

HARDIN, "Integrating the digitization of visual resources into library" in *Journal of Library Administration*; 39 (1) 2003: 45-55.

HARDY, OPPEHHEIM, RUBBERT, "Pricing strategies and models for the provision of digitized texts in higher education" in *Journal of Information Science*; 28 (2) 2002: 97-110.

LINDA HILL, GREG JANÉE, *et al.* "Collection Metadata Solutions for Digital Library Applications" in *Journal of American Society for Information Science*, 50 (13) 1999: 1169 – 1181.

ISAD (G) : general international standard archival description : adopted by the Committee on Descriptive Standards, Stockholm, Sweden, 19-22 September 1999. <www.ica.org/biblio/cds/isad_g_2e.pdf>.

RONALD JANTZ, "Public opinion polls and digital preservation: an application of the Fedora Digital Object Repository System" in *D-Lib Magazine*; 9 (11) 2003.
<<http://www.dlib.org/dlib/november03/jantz/11jantz.html>>.

ANNE KENNEY, OYA RIEGER, *Moving theory into practice: Digital Imaging for Libraries and Archives*, Research Libraries group, Mountain View, c2000.

KIESLING, (a cura di), Tag Library Conventions,
<<http://www.loc.gov/ead/tglib/tlc.html>>.

GRAHAM KLYNE, JEREMY CARROLL, "Resource Description Framework (RDF): Concepts and Abstract Syntax", <<http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/>>.

BRIAN LAVOIE, LORCAN DEMPSEY, "Thirteen ways of looking at... digital preservation" in *D-lib magazine*; 10 (7/8) 2004

STEVE LEE, „Writing for the Web – How to Write Web-Friendly Content for your Site" in <<http://www.freepint.com/issues/050603.htm#feature>>.

WILLIAM LE FURGY, "Levels of service for digital repositories" in *D-Lib Magazine*; 8 (5) 2002.

HENDRICK LENSCH, MICHAEL GOESELE, HANS PETER SEIDEL, "Digital collections of real world objects" in *D-Lib Magazine*; 8 (2) 2002.

Library of Congress, "Digital Library Standards"
<<http://lcweb.loc.gov/standards/metadata.html>>.

NORBERT LOSSAU, "Search engines technology and digital libraries" in *D-lib magazine*; 10 (6) 2004.

GABRIELE LUNATI, "Imago" in *Biblioteche oggi*, (8) 2002.

MAROSO, "Digitization as a preservation practice: look before you leap" in *Serials Librarian*; 45 (2) 2003:27-37.

TONY MC KINLEY, *From Paper to Web: How Make Information Instantly Accessible*, USA, Adobe Press, 1996.

MINERVA WORKING GROUP, "Good Practice Handbook".
<http://www.minervaeurope.org/structure/workinggroups/goodpract/document/goodpractices1_3.pdf>

MINERVA WORKING GROUP, Linee guida sulla costruzione dei siti culturali, novembre 2003
<<http://www.minervaeurope.org/structure/workinggroups/userneeds/document/s/cwqp-i.htm>>.

MINISTERO PER I BENI E LE ATTIVIT CULTURALI, ISTITUTO CENTRALE PER IL CATALOGO E LA DOCUMENTAZIONE, Normativa per l'acquisizione digitale delle immagini fotografiche, 2004
<<http://www.iccd.beniculturali.it/download/fotodig.pdf>>.

JACOB NIELSEN, *Web usability*, Milano, Apogeo, 2000.

BILL OLDROYD, THEO VAN VEEN, "Search and retrieval in the European Library: a new approach" in *D-Lib Magazine*; 10 (2) 2004.

GIOVANNI PASCUZZI, ROBERTO CASO, *I diritti sulle opere digitali: copyright statunitense e diritto d'autore italiano*, Padova, CEDAM, 2002.

MARK PEARROW, *Web usability: navigare, creare, gestire la rete*, Milano, Futura, 2001.

DAVID PEARSON, "Digitisation: do we have a strategy?" in *Ariadne*; (30) 2002. <<http://www.ariadne.ac.uk/issue30/digilib/intro.html>>.

PATRICK PECCATTE, Métadonnées: une initiation. Dublin Core, IPTC, EXIF, RDF, XMP, etc., 2004, <<http://peccatte.karefil.com/Software/Metadata.htm>>.

DANIEL PITTI, "Encoded Archival Description: An Introduction and Overview", in *D-Lib Magazine*, 5 (11) 1999.

PROCHASKA, "Special collections in an international perspective" in *Library Trends*; 52 (1) 2003: 138-150.

RANDALL, (a cura di), Development of the Encoded Archival Description DTD, revisione dicembre 2002 <<http://www.loc.gov/ead/eaddev.html>>.

RANDALL, (a cura di), Design Principles for Enhancements to EAD, revisione dicembre 2002 <<http://www.loc.gov/ead/eaddesign.html>>.

RICCARDO RIDI, "La biblioteca digitale: definizioni, ingredienti e problematiche" in *Bollettino AIB*, 44 (3) 2004: 273 – 344.

CHRIS RUSBRIDGE, "Towards the Hybrid Library" in *D-lib Magazine*, 4 (7) 1998, <<http://www.dlib.org/dlib/july98/07rusbridge.html>>.

ALBERTO SALARELLI, ANNAMARIA TAMMARO, *La biblioteca digitale*, Milano, Bibliografica, 2000.

SHELBY SANET, "The Cost to Preserve Authentic Electronic Records in Perpetuity: Comparing Costs across Cost Models and Cost Frameworks", in *RLG DigiNews*, October 2002 <<http://www.rlg.org/preserv/diginews/diginews6-5.html#feature2>>.

ANA SANLLORENTI, Digitalizacion y preservation digital, 2001.

MICHELE SANTORO, "Dall'analogico al digitale: la conservazione dei supporti non cartacei", *Biblioteche oggi*, 3 (19) 2001: 88 – 100.

MICHELE SANTORO, "Digitalizzazione dei documenti e conservazione del digitale", corso Ebsco Training Center, Milano, 13-14 novembre 2003.

MICHELE SANTORO, "Digitalizzazione dei documenti e conservazione del digitale"(parte II), corso Ebsco Training Center, Milano, 13-14 novembre 2003.

MICHELE SANTORO, "Territori digitali" in *Biblioteche oggi*, (8) 2002.

ANTONIO SCOLARI, *UNIMARC*, Milano, Bibliografica, 2000.

ORESTE SIGNORE, "RDF per la rappresentazione della conoscenza".

<<http://www.w3c.it/papers/RDF.pdf>>

ABBY SMITH, "Preservation in the future tense", in *CLIR issues*, (1998),3.

<<http://www.clir.org/pubs/issues/issues03.html>>.

ABBY SMITH, "Why digitalize?", <<http://clir.org/pubs/reports/pubs80-smith/pubs80.html>>.

ANNA MARIA TAMMARO, "Misurazione e valutazione della biblioteca digitale" in *Biblioteche oggi*, 1 2000: 66-69.

TECHNICAL ADVISORY SERVICE FOR IMAGES "An introduction to Digital Preservation" in *Technical Advisory Service for image , Advice Paper*

<<http://www.tasi.ac.uk/advice/creating/pdf/digipres.pdf>>.

TECHNICAL ADVISORY SERVICE FOR IMAGES, "Basic Guidelines for Image Capture and Optimisation" in *Technical Advisory Service for image , Advice Paper* <<http://www.tasi.ac.uk/advice/creating/pdf/img-capt.pdf>>.

TECHNICAL ADVISORY SERVICE FOR IMAGES, "Establishing a Digital Preservation Strategy " in *Technical Advisory Service for image , Advice Paper*

<<http://www.tasi.ac.uk/advice/creating/pdf/digipres2.pdf>>.

TECHNICAL ADVISORY SERVICE FOR IMAGES, "File Formats and Compression" in *Technical Advisory Service for image , Advice Paper*

<<http://www.tasi.ac.uk/advice/creating/pdf/formats.pdf>>.

TECHNICAL ADVISORY SERVICE FOR IMAGES "The Digital Image" in *Technical Advisory Service for image , Advice Paper*

<<http://www.tasi.ac.uk/advice/creating/pdf/image.pdf>>.

TECHNICAL ADVISORY SERVICE FOR IMAGES "Image Management Systems " in *Technical Advisory Service for image - Advice Paper*

<<http://www.tasi.ac.uk/advice/creating/pdf/ims2.pdf>>.

TECHNICAL ADVISORY SERVICE FOR IMAGES "Metadata and Digital Images" in *Technical Advisory Service for image , Advice Paper*

<<http://www.tasi.ac.uk/advice/creating/pdf/metadata1.pdf>>.

TECHNICAL ADVISORY SERVICE FOR IMAGES "Searching for and Retrieving Digital Images" in *Technical Advisory Service for Images – Advice Paper* <<http://www.tasi.ac.uk/advice/creating/pdf/srcandret.pdf>>.

TECHNICAL ADVISORY SERVICE FOR IMAGES "Setting up a Workspace for Digitization" in *Technical Advisory Service for image , Advice Paper* <<http://www.tasi.ac.uk/advice/creating/pdf/workspace.pdf>>.

TECHNICAL ADVISORY SERVICE FOR IMAGES "Scanners" in *Technical Advisory Service for image , Advice Paper* <<http://www.tasi.ac.uk/advice/creating/pdf/scanner.pdf>>.

TECHNICAL ADVISORY SERVICE FOR IMAGES "Sustainability of Digital Resources" in *Technical Advisory Service for Images – Advice Paper* <<http://www.tasi.ac.uk/advice/creating/pdf/sust.pdf>>.

TECHNICAL ADVISORY SERVICE FOR IMAGES, "Why "Archive Standards" " in *Technical Advisory Service for image , Advice Paper*. <<http://www.tasi.ac.uk/advice/creating/pdf/why-archive.pdf>>.

ROY TENNANT, "Metadata as if libraries depended on it" in *Library Journal*; 127 (7) 15 Apr 2002, pp.32, 34.

Testo della Legge 22 maggio 2004, n. 128. Aggiornamento della Legge 22 aprile 1941, n. 633. – Protezione del diritto d'autore e di altri diritti connessi al suo esercizio.

Testo della Legge 7 marzo 2001, n. 62 – Nuove norme sull'editoria e sui prodotti editoriali e modifiche alla legge del 5 agosto 1981, n. 416.

Testo del Decreto Legislativo 9 aprile 2003, n. 68 – Attuazione della direttiva 2001\29\CE sull'armonizzazione di taluni aspetti del diritto d'autore e dei diritti connessi nella società dell'informazione.

STEFANO VITALI, "Archivi on-line: qualche riflessione metodologica", <<http://www.storia.unina.it/perfez/vitali.rtf>>.

STEFANO VITALI, *Passato digitale: Le fonti dello storico nell'era del computer*, Milano, Bruno Mondadori, 2004.

IAN WINKWORTH, "Acquisire visibilità" in *Biblioteche oggi*, 6 2001:50-54.

WOODS, "When preserving the past - think of the future" in *Multimedia Information and Technology*; 28 (3) 2002: 71-2.

WORLD INTELLECTUAL PROPERTY ORGANIZATION, "Intellectual Property on the Internet: a Survey Issue", dicembre 2002. <<http://ecommerce.wipo.int>>.

World Wide Web Consortium, "Introduction to RDF Metadata",
<<http://www.w3.org/TR/NOTE-rdf-simple-intro>>.